DCRT

Division of Computer Research and Technology

National Institutes of Health

1993 Director's Report

U.S. Department of Health and Human Services
Public Health Service
Bethesda, Maryland
October 1993

Front and back covers: Carboxymyoglobin (MbCO) surrounded by 350 water molecules (P. J. Steinbach and B. R. Brooks, Molecular Graphics and Simulation Section, DCRT Laboratory of Structural Biology; see Proc. Natl. Acad. Sci. USA 1993; 90:9135-39). This study simulated the molecular dynamics of MbCO at 14 hydration levels, from 0 to 3,830 waters/protein. On the 100-ps time scale, 350 water molecules were found to fully hydrate the protein, covering all charged groups and resulting in an equilibrium structure and dynamics comparable to hydration by 3,830 water molecules.

Contents

Director's Preface

The past year has been an exciting and productive one for the Division of Computer Research and Technology. After 2 years of preparation – with site visits, retreats, consultation with outside advisors, and development of a reorganization plan – the implementation of that reorganization is now well under way. The past year has seen the creation of two new offices: the *Office of Computational Biosciences*, under my direction as an Acting Associate Director (equivalent to the Scientific Director) and an *Office of Computing Resources and Services*, headed by Emmett Ward as Acting Associate Director. The OCB consists of two new laboratories: the *Laboratory of Structural Biology* (Dr. V. Adrian Parsegian, Acting Chief) and the *Computational Bioscience and Engineering Laboratory* (Dr. Robert Martino, Chief), along with the pre-existing *Physical Sciences Laboratory* (Dr. George Weiss, Chief). The OCRS consists of two new branches: the *Networking Systems Branch* (Harold Ostrow, Chief), and the *Customer Services Branch* (Dale Spangenberg, Chief). In addition, the Personal Computing Branch has been restructured as the *Distributed Systems Branch* (David Songco, Chief), the Computer Center Branch has been restructured as the *Computing Facilities Branch* (Perry Plexico, Acting Chief), and the Data Management Branch has become the *Information Systems Branch* (Marvin Katz, Acting Chief). Ray Danner heads the *Statistical Support Staff*, and an *Architectural Management Staff* and a *Funding Management Staff* have been created. These restructurings are more than a change of name or a "facelift"; these are real changes in mission, orientation, strategy, tactics, personnel, physical location and resources. In addition to the OCB and OCRS, we have created a DCRT *Office of Information Resources Management*, headed by Arthur Schultz; Frances Halverson has been named Assistant Director for Programs; and the *Capacity Management Staff* continues to play its important function.

The reorganization has been accomplished without an addition to our resources in terms of budget or personnel. This has necessitated the closing of three laboratories within the division. We believe that the majority of the restructuring is now complete, and that the current structure of DCRT should be appropriate for several years to come.

Further details about the reorganization are given on pp. 11-14.

Major accomplishments during the past year include:

- reorganization of the entire division
- the application of High Performance Computing and Communication (HPCC) to several important biomedical problems
- porting of the programs CHARMM (molecular dynamics) and GAMESS (quantum mechanics) to the Intel® highly parallel supercomputer and to Hewlett-Packard™ workstation clusters
- the DCRT cosponsored meetings, "High Performance Computing in Chemistry" and "Intelligent Systems in Molecular Biology"
- improved measurement of forces between molecules
- improved understanding of the role of water surrounding proteins in allosteric transitions, protein conformation, computational molecular dynamics, and simulated *de novo* folding of proteins
- improved prediction of protein secondary structure
- classification of protein structures
- a new model for assembly of clathrin-coated vesicles
- expansion of NIHnet to include 224 local area networks (LANs) with 105 on-campus LANs connected via high-speed fiber backbone
- provision of new information services over the network, including MEDLINE, Current Contents®, REFERENCE UPDATE®, *Gopher™*, and multiple genetic databases such as GenBank
- a new set of software tools for medical and laboratory image processing (Multimodality Research Image Processing System (MRIPS), with LDRR (NIH/OD), NCRR and the ICDs)
- a successful first year of operation for the Scientific Computing Resource Center

- presentation of the 1992 Best of Open Systems Solutions (BOSS) Award for Innovation in Hardware, Software and Networking Approaches to DCRT's Advanced Laboratory Workstation (ALW) Project
- new courses on molecular modeling and experimental design
- support for large-scale sequencing efforts (with NINDS and NCI)
- the evaluation of client/server technologies
- major efforts on such information resources management activities as disaster recovery, data security, new procurement mechanisms and site licensing
- advances in hardware and technology (e.g., workstation clusters)
- participation in NIH-wide activities, including training, lectures, seminars, and journal clubs; organizational consults (e.g., NCHGR, NINR); and the development of a new User Resource Center at Executive Plaza (with the Division of Personnel Management)
- support for NIH Office of the Director administrative systems
- reconfiguration of the Central Computer Facility
- substantial rate reductions and rebates for mainframe services
- improved liaison with the user community
- planning and initial studies for reprocurement of the mainframe, and appointment of a "trailboss" to manage the reprocurement and coordination with multiple Federal agencies
- upgrade of the speed of telecommunications interfaces
- new equal employment opportunity initiatives
- adoption of a high school. to encourage students to enter careers in computing and biomedical research.

Beyond these present accomplishments lies the future of DCRT, and the scientific and administrative computing advances needed for NIH to move into the twenty-first century.

David Rodbard, M.D.
Director, DCRT

Office of Computational Biosciences

To exploit the power of the division's parallel scalable supercomputer (the Intel® IPSC i860 128-node machine), the *Computational Bioscience and Engineering Laboratory* has made significant progress in the adaptation and parallelization of several computer codes for important biomedical applications. Applications include:

- image processing
- computational chemistry
- quantum chemistry
- genetic database searching
- protein structure determination
- protein structure prediction
- multiple sequence alignment.

A brief summary is shown in Table 1. This work has involved collaborations with investigators in NIAMS, NIDDK, and NCI, and in DCRT's Laboratory of Structural Biology. For practical purposes, every application which has been attempted has been successful. In each case, we are seeing speedups in processing time of 50- to 100-fold, close to the theoretical maximum, making it possible to attempt projects which otherwise would have been impossible or impractical. This is the major effort at NIH in the area of High Performance Computing Systems, part of the High Performance Computing, Communications, and Information Technology initiative of the Federal Coordinating Council for Science, Engineering, and Technology of the Committee on Physical, Mathematical, and Engineering Sciences. In this regard, we cosponsored a meeting on "High Performance Computing in Chemistry" jointly with the Pacific Northwest Laboratories of the Department of Energy. Dr. Robert Martino and coworkers were also important contributors to a meeting sponsored by NASA on high performance computing, and the division was a cosponsor of a meeting on Intelligent Systems in

Table 1. DCRT High Performance Biomedical Computing Program
Representative Collaborative Research Activities

Biomedical Application	*Collaborator*	*Organization*
Structural Biology		
Electron Microscopy	A. C. Steven	LSBR, NIAMS
NMR Spectroscopy	A. Bax	LCB, NIDDK
x-Ray Crystallography	C. G. Hyde	LSBR, NIAMS
Protein Folding Prediction	B. Lee	LMB, NCI
Medical Image Processing		
PET Reconstruction	R. E. Carson	NMD, CC
Functional Neurological Analysis	J. V. Haxby T. Zeffiro	LPP, NIMH LN, NIA
Computational Chemistry		
Quantum Chemistry	B. Hardy	DBB, CBER
Molecular Dynamics Simulations	W. A. Eaton E. Henry	LCP, NIDDK LCP, NIDDK
Genetic Linkage Analysis	E. S. Gershon	CNB, NIMH
Radiation Treatment Planning	J. van de Geijn	DCT, NCI

Molecular Biology immediately preceding the national meetings of the American Association for Artificial Intelligence. Our Intel® parallel supercomputer is now being utilized at capacity, and it is imperative that we obtain an appropriate follow-on machine with additional capacity to keep us current in terms of the rapidly evolving technology and increasing applications.

Dr. Benes Trus, Chief of the Image Processing Section, CBEL, as made considerable progress in his studies of the structure of the herpes simplex type 1 virus. Using computer analysis of cryoelectron micrographs obtained by his colleagues and collaborators in the Laboratory of Structural Biology Research of NIAMS, he has been able to make tentative identification of the role of each of the seven proteins comprising the capsid. The studies have been facilitated by the ability to selectively extract one or another of the proteins, leaving a hole where it once was, and by the use of monoclonal antibodies selective for each of the proteins (Figure 1). The investigations are contributing to an improved understanding of the structure (and hence function) of the various parts of the virus capsid (e.g., "pentons," "hexons," VP1, VP2, VP7). This should facilitate studies of the mechanisms of infectivity and replication of the virus, and lead to the development of pharmacological agents to interfere

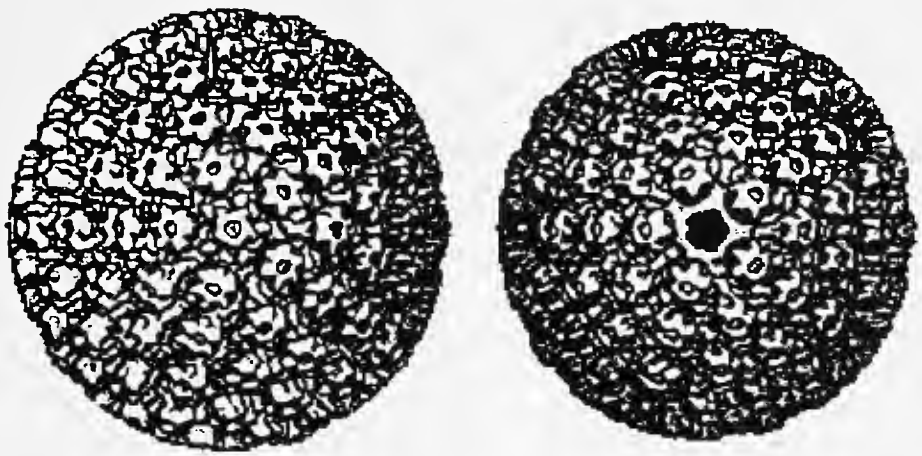


Figure 1. The Computational Bioscience and Engineering Laboratory, DCRT, in collaboration with the Laboratory of Structural Biology, NIAMS and the University of Virginia, Charlottesville, is involved in the 3D reconstruction of viruses. Shown are surface-shaded representations of herpes simplex virus (HSV) type 1. The HSV-1 capsid has a diameter of ~125nm, and is constructed according to icosahedral symmetry. The left image is an empty capsid as viewed down the five-fold axis of symmetry. The locations of the various capsid proteins have been investigated by biochemical depletion experiments. Treatment of capsids with the denaturant, guanidinium hydrochloride, extracts certain capsid components while preserving its icosahedral geometry, as shown in the right image. The pentons are removed quantitatively; cf. the empty vertex sites. Also removed are some of the triplexes — nodules that occupy the local sites of three-fold symmetry. (Electron microscopy, image reconstruction, and computer graphics: B. L. Trus (DCRT and NIAMS) and F. P. Booy, J. F. Conway, and A. C. Steven (NIAMS); virus preparation, biochemistry: J. C. Brown and W. W. Newcomb (University of Virginia); reconstruction software: T. S. Baker (Purdue University) and C. A. Johnson and N. I. Weisenfeld (DCRT). Computationally intensive steps were performed on the DCRT Intel iPSC/860 parallel supercomputer.

with the life cycle of this virus. The studies should generalize to a variety of other viruses, both in the herpes family and others. Dr. Trus' work also illustrates the importance of high performance (read, "parallel scalable") computing in facilitating and accelerating biomedical research.

In the *Laboratory of Structural Biology*, Dr. Sergey Leikin has measured the forces involved in the packing of collagen triple helices, as a function of temperature, pH, ionic strength and ionic milieu, using the "Parsegian" method to apply osmotic force and x-ray diffraction to measure intermolecular distance. Surprisingly, hydration forces are a major determinant of these forces, as in the case of DNA-DNA, DNA-protein, and lipid-lipid interactions. These results will now be examined in a variety of disease conditions involving abnormal collagen structures. Dr. Leikin and Dr. Parsegian continue a number of important collaborations with researchers in NIDDK and NIAMS.

Dr. Bernard Brooks (Chief, Molecular Graphics and Simulation Section) and Visiting Fellow Dr. Milan Hodoscek have adapted the CHARMM program for molecular dynamics to the Intel® parallel computer, and in turn, to a cluster of workstations, providing economical and cost-effective methods for long-term simulations. Dr. Peter Steinbach has examined the role of hydration in modulating the structure of myoglobin by systematically varying the degree of hydration from zero to a high level (see figures on front and back covers of this report). Dr. David Chatfield is making significant progress in the modeling of alternative reaction mechanisms for HIV protease, using a combined quantum mechanical/molecular mechanics (QM/MM) approach.

Dr. Peter Munson (Chief, Analytical Biostatistics Section), Dr. Raul Porrelli and Dr. Valentina di Francesco (Visiting Fellows) have used a variety of sophisticated statistical methods to predict the secondary structure of proteins on the basis of primary sequence. By use of generalized cross-validation, they have been able to compare the performance of different methods which explicitly or implicitly use different numbers of parameters. This rigorous statistical method helps to make sense of the conflicting claims in the literature, and show that all methods are reaching an upper plateau of about 62% accuracy. Further improvement will require additional information, e.g., prior classification of proteins into various categories.

Richard Feldmann is exploring a variety of new and novel approaches to analysis of the "topology" of *de novo* protein folding.

Dr. George Weiss, Chief of the *Physical Sciences Laboratory*, has completed a major monograph entitled "Introduction to Crystallographic Statistics" and edited a volume entitled "Contemporary Problems in Statistical Physics." He has continued work with Dr. Uri Shmueli (Visiting Scientist) on the theory of crystallography of small molecules.

Dr. Ralph Nossal and Dr. Albert Jin (Visiting Fellow) have developed a new and novel theory for the mechanism of assembly of clathrin triskelions to form coated vesicles. This model appears to solve several serious geometrical and physical problems, and is now ready for experimental evaluation.

In addition to our seminar series, DCRT sponsors a journal club devoted to protein folding, and another devoted to biomedical applications of artificial neural nets.

Finally, mention should be made of DCRT's Image Technology Program, headed by the Clinical Center's Dr. Stephen Bacharach under a joint DCRT/CC appointment. This program links DCRT staff to collaborators in the CC and several ICDs with the goal of developing improved methods for *in vivo* image analysis and processing. Current projects in which DCRT is involved include:

- 3D alignment of positron emission tomography (PET) transmission scans
- automatic tracking of magnetic resonance imaging (MRI) "Tag" grids
- maximum likelihood estimation of regional radioactivity concentration
- computer-guided surgery.

David Rodbard, M.D.
Acting Associate Director, OCB

Office of Computing Resources and Services

DCRT finalized a Strategic Plan and reorganized to align ourselves with that plan during FY93. The plan and thus the new organization address three major programs: Research and Development, Computer Resources Infrastructure, and Direct Computing Services and Support. The Director, NIH approved the establishment of two new offices, one dealing with services, support, infrastructure and facilities (Office of Computing Resources and Services (OCRS)), and the other with research and development for the computational biosciences (Office of Computational Biosciences (OCB)). Key elements of the OCRS organization include:

- creation of a central point of contact for all services and support in the division
- a central focus for campus-wide networking
- consultation and support for evolving distributed systems technology
- consolidated operation, maintenance and support of all DCRT hardware and software platforms for shared and enterprise use
- creation of a core mechanism for identifying and evaluating opportunities to make the transition to open systems environments.

The functional definition of the new branches of the OCRS and the reallocation of resources is well under way at this writing. A quick list of the branches and their primary functions follows:

- *Network Systems Branch* (NSB). Designs, develops and supports all network facilities and services related to NIHnet, the NIH-wide backbone infrastructure; fosters computational interoperability; and promotes the development of state-of-the-art networking technology.
- *Computing Facilities Branch* (CFB). Develops, operates, maintains and supports central hardware and software platforms for shared and enterprise use; evaluates, installs and maintains central servers, gateways and database management facilities that support client/server computing; and takes the lead in devising and implementing workable strategies for migrating NIH computing to open systems.
- *Distributed Systems Branch* (DSB). Addresses the increasing demand for service, support and guidance in the selection and effective use of personal computers, workstations, local area networks, and associated automation technology; provides advice and assistance on issues relating to multiplatform client/server and database support; and provides primary planning and support for the Scientific Computing Resource Center (SCRC).
- *Customer Services Branch* (CSB). Serves as the primary user contact for information, support and training within DCRT; manages and facilitates the resolution of user problems with the appropriate staff in DCRT; and, in general, acts as the user advocate within the DCRT.
- *Information Systems Branch* (ISB). Provides continued support for the NIH Administrative Data Base (ADB), the Central Accounting System (CAS), and the Clinical Information Utility (CIU); serves as an NIH resource for database design, systems analysis and programming; and plays the lead role in evaluating, selecting and supporting NIH client and local workstation and server database products.

In addition to the branches, a Statistical Support Staff has been established to provide direct advice, assistance and support to biostatisticians and others at the NIH who either are using or are planning to use statistical software on central and distributed platforms. Representatives from each of the branches also participate on the Architectural Management Staff and the Funding Management Staff. The objectives of these two staff groups are, respectively, to foster collaboration related to DCRT-wide architectural planning, and to identify and develop mechanisms for new DCRT cost recovery alternatives.

The Customer Services Branch is already preparing to assume its pivotal role in the reorganization. A client/server facility is being developed to support problem tracking and resolution across the OCRS; plans for a new centralized training program are in place. Consolidation of user services with a central point of contact and a single

phone number for assistance will help to speed the NIH researcher or administrator to the proper DCRT resource for his/her information and support requirements.

Even as plans to move support for open systems to the CFB were in gestation, the Federal Computer Conference bestowed an honor upon the staff of the Advanced Laboratory Workstation (ALW) Project, formerly part of the Computer Systems Laboratory. The ALW system received the 1992 Best of Open Systems Solutions (BOSS) Award for Innovation in Hardware, Software and Networking Approaches.

Higher communications speeds and enhanced error correction for the CFB's interactive services – in the form of new communication controllers and new state-of-the-art modems – will open up capabilities and functions such as large file transfers, which have not been previously viable.

For the 25th consecutive year, cost savings were passed on to users of the computer center in the form of significant rate reductions, rebates, and discounts ranging from 21% to 28%.

The Scientific Computing Resource Center (SCRC), now located in the DSB, was piloted in May 1992 and is now flying strongly, especially with the opening of its Image Technology Center in July 1993. The center has been particularly popular for molecular modeling, sequence analysis, graphics and statistics.

DSB and other branches have collaborated in beta-testing new products such as Windows™ NT and the various Lotus 1-2-3 releases. This positions DCRT to influence product enhancements that will meet the particular requirements of the NIH community.

DSB's Dr. Dale Graham has developed new courses and training manuals to assist in the use of GenBank, other databases, MacVector™, GCG and other sequence analysis programs. John Powell continues to play a major role in the automation of laboratories performing large-scale sequencing. His expertise includes hardware, software, networking and databases, engineering, and application of new technologies such as the "fast data finder" chip and the commercial Inherit™ system. He has recently been joined in this effort by Dr. Mark Miller from NCI, and is providing major support to researchers in NINDS, NCI and NCHGR.

In addition, DSB's Dr. James Malley has completed a monograph on "Quantum Statistical Inference" which is now in press as a series of journal articles.

The Statistical Support Staff sponsored a Mathematical and Statistical Software Fair at NIH during December 1992. This was the first of its kind, and it introduced NIH mathematicians and statisticians to multiplatform mathematical and statistical packages. A questionnaire was distributed among the attendees and important data were compiled on software packages of interest to the NIH community.

Interest in the ADB Information System (ADBIS) was such that several formal demonstrations were presented in the Lipsett Amphitheater. The ADBIS represents the fruition of a collaborative effort among ISB staff and over 70 representatives from all of the ICDs. This effort is being coordinated by Mr. Mark Kochevar of NCI who is serving as Chairman of this ADB Steering Subcommittee. The ADBIS is an online system which provides standard query facilities that are specifically designed to respond to the requirements developed by the subcommittee.

During the fiscal year, the increased speed of Fiber Distributed Data Interface (FDDI) was extended to an additional 110 local area networks (LANs) in 10 buildings on the NIH campus. FDDI operates at 100 megabits per second and portends the ability to accommodate transfer of large files for image processing, full-motion video, genome mapping and other research applications that require massive data transfer at high speed. Currently, there are about 250 LANs on the NIHnet, which serves sites on and off the campus. This number will probably grow to around 330 during the coming year, and NSB plans to provide the most advanced, appropriate and latest supportable technology for each site.

Valuable services have been provided to the NIH community as a whole. For example, the PUBnet Fax gateway allows any LAN user at NIH to send electronic faxes to anywhere in the world; antiviral software for PCs has been made available to virtually all NIH employees; and many electronic NIH forms have been made available for down-loading from PUBnet in the various formats most used by the NIH community.

A new Macintosh® database and desktop publishing system for producing the NIH Scientific Directory/Annual Bibliography (SD/AB) simplified ICD submission requirements, easing the pain for ICD coordinators and facilitating the job of the Editorial Operations Branch, OD in producing this year's SD/AB book.

If it were possible to accurately describe and estimate the true costs of a distributed computing investment, we might be able not only to make more informed systems design decisions, but also to identify clear opportunities to reduce costs and optimize resource requirements. To do this, one must consider the costs and resource commitments that go beyond initial hardware and software purchases. These costs include support, training, system administration, backup and recovery, and hardware and software upgrades. It has been estimated that these follow-on costs represent three to four times those of the initial purchase. DSB is actively pursuing an independent and objective analysis of these costs with the Gartner Group, Inc.

Strategies and Plans for the Future

The OCRS is aggressively pursuing a number of initiatives designed to better serve the NIH community in a world of rapid technological change.

In the networking arena, several initiatives are in progress and others are planned. Construction of the NIHnet backbone and consolidation of RESnet, NUnet, and CCnet into a cohesive whole has been completed. Value-added information resources and network-based applications are now being developed rapidly – by DCRT, the ICDs, and by academia and industry nationwide. This sets the stage for distributed computing and realizing the benefits of client/server technology. A Microsoft® Mail gateway is in "pilot production" and currently handles mail from 20 LANS at NIH. When complete, NIH will have a cross-platform, transorganizational communication system which includes e-mail directory synchronization among servers, user address exchange with the NIH e-mail directory, backup and recovery, and operational monitoring.

The initial model of an NIH-wide mail directory is in a test stage, and is scheduled to be made generally available to the community early in FY94. When complete, the mail directory will provide transparent access to addresses of all NIH electronic mail users. As Microsoft® Mail enters full production, we hope to fully integrate its directory services with the NIH-wide mail directory.

As part of our quest for additional value-added services on the network, we are examining the feasibility and potential cost benefits of expanded site licensing for, and network distribution of, commonly used LAN and desktop software. Negotiating site licenses for the campus would make it simpler and cheaper for NIH to obtain software and related upgrades for facilities such as heavily used operating systems, word processors, desktop client and run-time modules, and statistical programs. Broad use of this concept could also reduce administrative costs related to procurement, and provide mechanisms across the NIH community and within the ICDs to better coordinate and control the proliferation of multiple versions of the same software.

Gopher™ is a network-based distributed information search and retrieval system. Developed at the University of Minnesota, *Gopher™* comprises both a protocol and client/server software, and provides access to a wide variety of information and network resources. DCRT introduced *Gopher™* at NIH in the summer of 1992 on the NIH Convex system, through combined efforts of the Convex staff and the Computational Molecular Biology Section.

Gopher™ is a truly revolutionary step towards making the Internet and its resources available to users, and over 1,500 sites around the world now provide a uniform, simple interface to an astounding volume and variety of information. At NIH, through the collaborative efforts of many, access has been provided to:

- Current Contents® and REFERENCE UPDATE®
- Molecular Biology databases including GenBank, PIR, SWISSPROT, Protein Data Bank, PROSITE, Listing of Molecular Biology Databases (LiMB), and Transcription Factor Database (TFD)
- NIH Phone Book and e-mail directory
- Current Index to Statistics
- NIH Guide to Grants and Contracts
- National Cancer Institute's CancerNet
- CRISP (Computer Retrieval of Information on Scientific Projects) System
- Catalogs of the DCRT and NIH libraries.

Several components of the OCRS have successfully developed prototypes of client/server applications and are addressing the many issues of cross-product connectivity which arise in an open systems environment as a barrier to interoperability. Investigation of products that might be used to establish an effective client/server environment at NIH is actively being pursued. During the coming year, DCRT plans to implement and fully support client/server gateways to its mainframe and central servers and to provide highly interoperable support for other processing platforms. We will also collaborate to select and support client software and LAN database products.

Probably the most visible change in DCRT will occur in the direct customer service area. In the past, DCRT has provided excellent user service for several of our highly visible functions. However, new and even regular users were often confused by DCRT's myriad of services and related contact points. Our plan calls for broadening the existing walk-in service provided by CFB to include all DCRT services, and providing a single, easy-to-remember phone number, i.e., 4-DCRT, for all remote inquiries. CSB plans to gradually transition existing services in a manner that ensures a reasonable evolution to one-stop customer service.

J. Emmett Ward
Acting Associate Director, OCRS

DCRT Reorganization

From June to December 1992, the Deputy Director and four senior managers slowly, carefully, and methodically developed a plan to reshape the division. The group sought to position DCRT so that it would be most relevant and responsive to the needs of the NIH scientific, administrative, and extramural communities. The group found, for example, that networking had previously been handled by two separate units within the division; database was being handled by three separate units. Users would have to deal with a dozen or more different contact points in the division. As a result, it was difficult for users to find the help they wanted and needed. The reorganization was designed to help correct these problems, to reduce or eliminate redundancy, to permit growth in high-priority areas, to become more efficient and cost effective, and to be in a position to change in response to changing demands from the NIH community and changing technology.

First, the division was split into two major components: the *Office of Computational Biosciences*, headed by the Scientific Director; and the *Office of Computing Resources and Services*, headed by a newly appointed Associate Director for Computing Resources and Services.

The Office of Computational Biosciences has three laboratories, two of which are newly created:

The *Computational Biosciences and Engineering Laboratory* is headed by Laboratory Chief Dr. Robert Martino. This laboratory is dedicated to applying the most advanced, high performance supercomputer technology to the problems of biology and medicine. It is already at the forefront with a 128-processor Intel® IPSC i860 supercomputer, which has been applied to computationally intensive problems such as

- molecular dynamics
- quantum mechanics
- protein structure determination by crystallography and multidimensional NMR spectroscopy
- clinical imaging, including the registration of images from PET, CT and MRI
- reconstruction of the 3D structure of viruses from cryoelectron microscopy (in collaboration with NIAMS)
- genetic database searching.

Dr. Benes Trus heads a section dedicated to biomedical image processing.

The *Laboratory of Structural Biology*, headed by Dr. V. Adrian Parsegian, brings together scientists with interests in protein structure and its prediction, molecular dynamics and modeling, and the measurement of forces between macromolecules (proteins, nucleic acids, polysaccharides) using laboratory studies (with longstanding support, collaboration and facilities from NIDDK) combined with sophisticated theoretical analyses. Section heads include Drs. Bernard Brooks, Peter Munson, and V. Adrian Parsegian.

The *Physical Sciences Laboratory*, headed by Dr. George Weiss, continues to apply methods of mathematical modeling and physics to biomedical problems, including the theory of crystallography and the study of biopolymers, fractals, and image processing (with the CC Department of Nuclear Medicine). Dr. Ralph Nossal heads a group interested in laser light scattering, biological imaging, and the organization of biological polymers such as clathrin-coated pits. Dr. Nossal conducts experimental studies in collaboration with NCRR, and neutron scattering studies at the National Institute of Standards and Technology.

The Office of Computing Resources and Services now comprises five branches and the SCRC. These branches provide consultation, service and support for a wide range of platforms and applications, including:

- PC and Macintosh® microcomputers
- UNIX® workstations and workstation clusters
- networking
- central support for distributed computing
- database design, development, operation, and maintenance
- mainframe and supercomputer services.

The five branches include:

• the *Network Systems Branch* (NSB, Harold Ostrow, Chief), formed by consolidating the talents of networking specialists formerly scattered throughout three DCRT components. NSB's consolidated mix of talents will help the branch provide the hardware and software infrastructure to allow NIH's many different computer systems to share information with each other and with national and international networks. A "hot line" as well as ongoing individual contacts via telephone, meetings and e-mail will form an NSB support system that will aid ICD Technical Local Area Network Coordinators (TLCs) and their users.

• the *Distributed Systems Branch* (DSB, David Songco, Chief). Evolving technologies caused the Personal Computing Branch to change its focus – and its name – to distributed systems. Distributed computing requires a new model that will provide specialized support for user-owned personal computers, workstations, local area networks, and the associated automation technology. The new branch will help customers with larger and more complex problems; a new focus will be on guidance for groups in the effective and efficient use of distributed computing. Developing solutions to NIH-specific problems is paramount; the branch wants the scientific and administrative communities at NIH to be able to focus on their research and their work, not on computers, and aims to help them do that.

• The big difference between the *Computing Facilities Branch* (CFB, Perry Plexico, Acting Chief) and the old Computer Center Branch is scope of operations. The branch will continue to manage the traditional computer center resources like the IBM® mainframe and Convex supercomputer, but will also take on the other centrally owned division resources, like the Advanced Laboratory Workstation project and the Intel® massively parallel supercomputer. New strategic directions are in store for the branch, with the biggest challenge being to combine the sometimes disparate resources offered to the community into a cohesive whole.

• The *Information Systems Branch* (ISB, Marvin Katz, Acting Chief), formerly the Data Management Branch, will continue to advise and serve NIH customers in developing and maintaining computer-based information systems. One particular challenge will be the re-engineering of "legacy" systems such as the Administrative Data Base System (ADB), carefully leveraging the investment NIH already has in its systems. Most of all, the branch wants to get users more involved in joint application development.

• The *Customer Services Branch* (CSB, Dale Spangenberg, Chief) reflects a renewed DCRT commitment to its customers. Included in this commitment will be a central point of contact – a single phone number to help NIH employees navigate DCRT's varied services – and training to address common needs for information. The branch believes in a consistent, centralized multiplatform approach to service and support.

• The *Scientific Computing Resource Center* (SCRC, Dr. Brian McLaughlin) provides the NIH with a shared-use facility, staffed by computer professionals, where researchers are able to focus on scientific computing. The SCRC is dedicated to addressing the needs of the NIH scientific community by providing access to scientific software running on advanced personal computers and UNIX® workstations. It makes available different types of scientific computing solutions, so that researchers can make informed decisions about which of these should be incorporated into the various ICD resources.

In addition, several other new entities were created:

• The *Assistant Director* (Frances Halverson) will address a variety of division administrative and policy issues.

• The *Office of Information Resources Management* (OIRM, Arthur Schultz, Chief) provides an important focus for DCRT's IRM issues, including evaluating available technologies, selecting existing contracts for procurements, investigating site licenses for software, and computer security. One critical task is to help upgrade DCRT mainframes and supercomputers to keep up with advancing technology and the needs of NIH.

• The *Architectural Management Staff* will track and evaluate rapidly evolving technologies and make decisions about which products to offer to the NIH community.

• The *Funding Management Staff* will identify and develop cost recovery mechanisms and propose ways to implement them.

• The *Statistical Support Staff* (Ray Danner, Head) provides support to the biostatistical community and to biomedical researchers. It consults in quantitative analysis and associated computer use, and selects, maintains, and supports mathematical and statistical software for mainframes, personal computers and workstations. It provides site licenses for several popular programs on several platforms, and offers an extensive training program.

In the process of reorganization, it has been necessary to phase out and close three laboratories (the former Laboratory of Applied Studies, the Laboratory of Statistical and Mathematical Methodology, and the Computer Systems Laboratory) and to transfer the resources and personnel from those laboratories to the new laboratories. In other cases, sections have been created or closed. The serious limitation on availability of funding and personnel has meant that almost all of the changes and new appointments had to be made internally. This has helped to provide some significant upward mobility for personnel within the division.

The new organizational structure is shown in Figure 2. This structure will allow us to be more responsive to the high priority areas of NIH and of DCRT, thus facilitating the graceful expansion of services and functions anticipated for the next 5 years. It is likely that the new structure will be suitable for the next several years; however, additional changes will be made as needed.

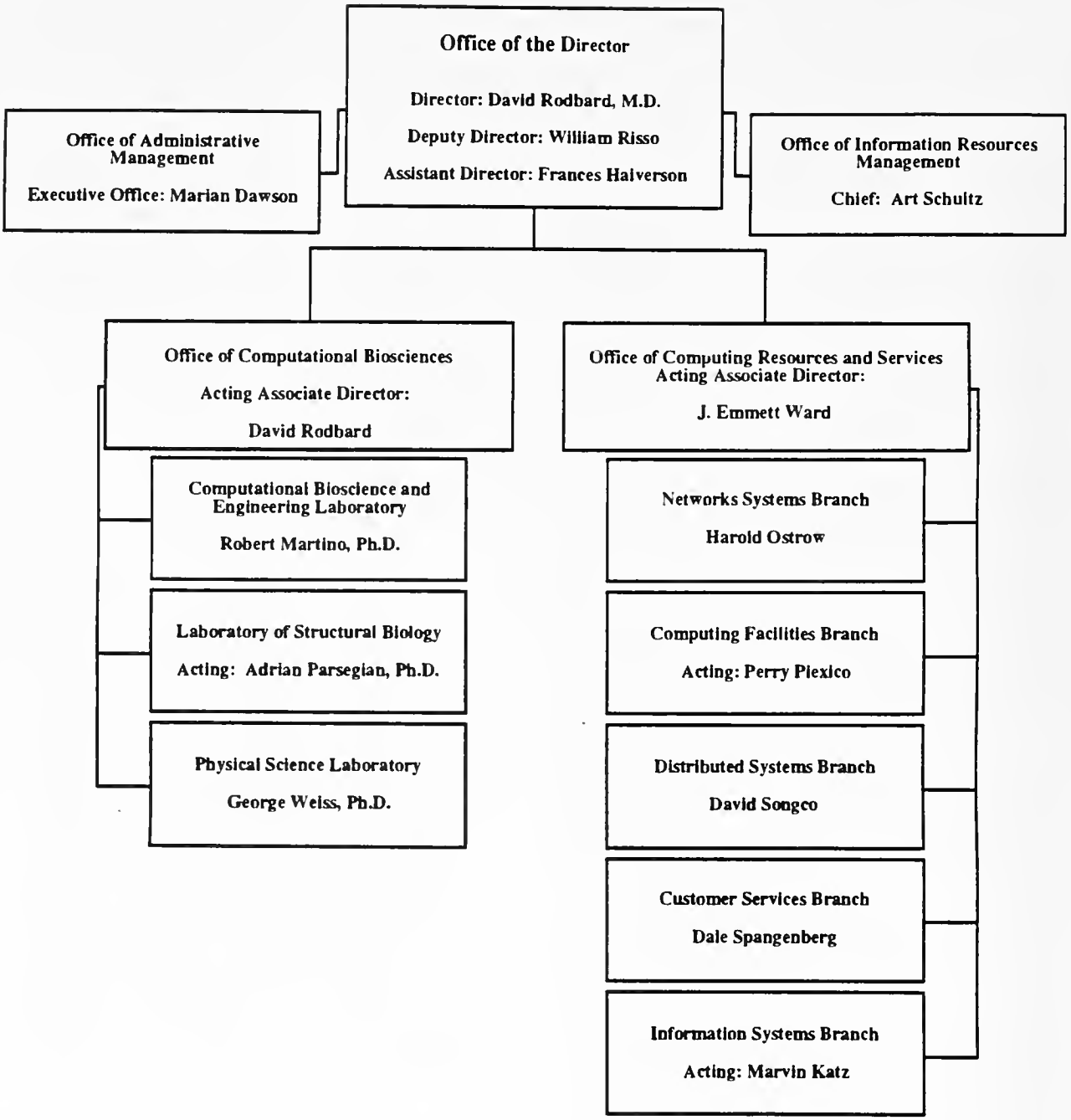
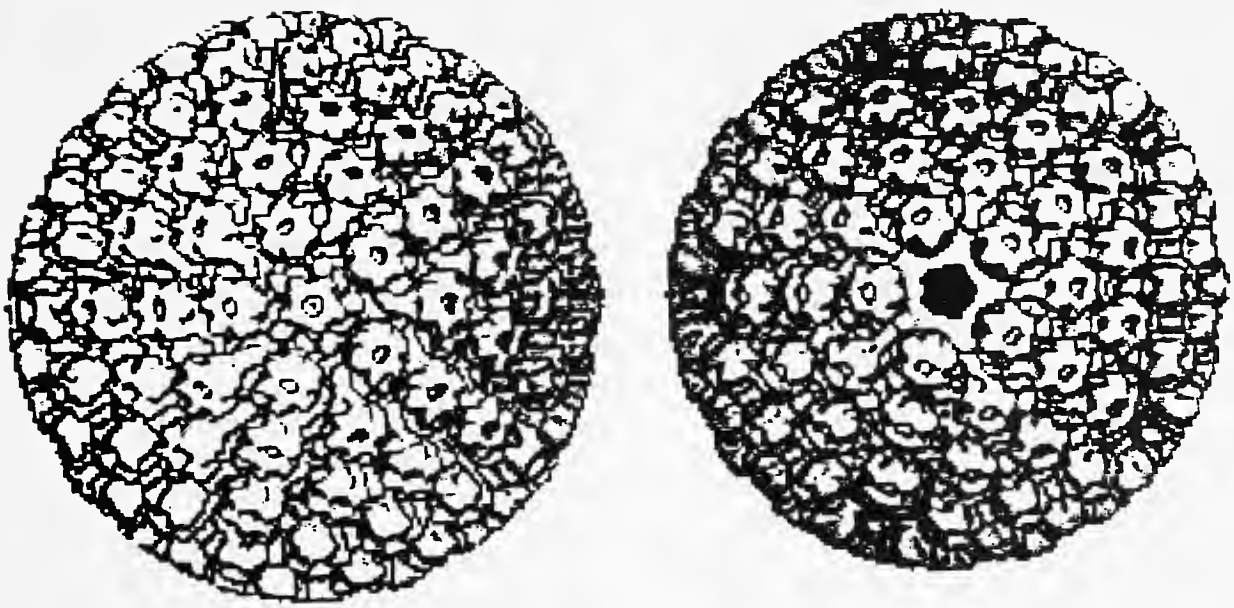


Figure 2. DCRT organizational structure

CBEL

Computational Bioscience and Engineering Laboratory



Computational Bioscience and Engineering Laboratory

Robert L. Martino, Ph.D., Chief

The Computational Bioscience and Engineering Laboratory (CBEL) is a newly established DCRT laboratory devoted to the exploitation of high performance computer systems in biomedical applications including image processing, structural biology, computational chemistry, medical imaging, scientific visualization, signal processing, genetic database searching, genetic linkage analysis, and advanced statistical methods. Members of CBEL strive to identify and solve those computational problems in biomedicine that can benefit from high performance hardware, modern software engineering principles, and new and efficient algorithms. The laboratory also provides high performance parallel computer and image processing systems for the NIH scientific staff.

For the future, DCRT recognizes the strategic importance of high performance biomedical computing. Eight of the twelve initiatives described in DCRT's long-range plan involve this activity:

- research in the emerging discipline of computational biosciences
- investigate applications of high performance computing in biomedical research
- establish an NIH shared-use digital image processing capability
- develop and enhance the methods of mathematical modeling for a new generation of scientific problems and computer technology
- research to facilitate biomedical data access and use
- foster distributed computing throughout the NIH administrative and research infrastructures
- enhance scientific computing and networking resources
- expand support and service for scientific computing.

CBEL was formed in recognition of the need to have a laboratory that would make a contribution to these important division activities through its efforts in applying new high performance system architectures and computer engineering principles to the solution of biomedical computing problems.

CBEL provides leadership in the research, development, and biomedical application of massively parallel computers in a networked environment. It collaborates with research investigators in modeling of complex systems, analyzing and interpreting data, signals and images, and assisting with computationally intensive tasks in application areas that include electron and light microscopy in the study of biological structure and function, x-ray crystallography and NMR spectroscopy for protein structure determination, molecular dynamics and quantum chemistry in the design of pharmaceuticals, medical imaging to study brain function, and radiation treatment planning for the treatment of cancer. CBEL conducts continuing research to expand the use of high performance computing in biomedical areas. It develops research systems into progressively more accessible and user-friendly systems which ultimately become routine computer utilities.

The work of CBEL staff has contributed to a number of findings of biomedical significance in the past year. Working with NIAMS collaborators, progress was made on determining the three-dimensional location of the seven major capsid proteins of the herpes simplex virus type 1. CBEL staff implemented a system to quantitate lens opacities that is sensitive enough to show cataract progression in one year. No commercially marketed instrument has been able to show this ability. NIDDK has used parallel computing methods to improve its procedure for determining the structure of the protein calmodulin from NMR spectra data. Another group of scientists from NIDDK simulated the kinetics of nitric oxide rebinding to myoglobin following photodissociation on the CBEL parallel computer. NIMH investigators used parallel image registration techniques developed by CBEL staff to study the progression of Alzheimer's disease from PET images

of the brain. High performance computing has allowed NEI researchers to determine the onset time, the rate of information encoding, and the total amount of information encoded by the neuronal responses to different parameters of a visual stimulus in primates.

CBEL executes its work through an Office of the Chief and two operating sections:

- the *Office of the Chief* provides overall CBEL management and planning including laboratory administrative and financial functions. It coordinates the establishment of new laboratory activities and the work of the sections to encourage and ensure appropriate cooperation and integration of effort. It also coordinates CBEL work with other parts of DCRT and the NIH ICDs as well as other government agencies and research institutions.
- the *High Performance Computing Section* (HPCS), with the CBEL Chief Dr. Robert Martino serving as Acting Chief, develops high performance computer systems for the solution of biomedical laboratory and clinical research problems. It provides parallel algorithm expertise to solve computationally intensive problems in biomedicine. HPCS deploys modern, nontraditional, computer architectures in a distributed computing environment and provides a high performance parallel computer facility for the NIH scientific staff.
- Dr. Benes Trus, who holds a joint appointment with the Laboratory of Structural Biology Research, NIAMS, is the Chief of the *Image Processing Research Section* (IPRS). IPRS creates and adapts algorithms and computational techniques for various biomedical imaging modalities including electron microscopy, light microscopy, Positron Emission Tomography (PET), Single Photon Emission Computer Tomography (SPECT), and Magnetic Resonance Imaging (MRI). It performs research in structural biology and biochemistry using state-of-the-art image processing methods. IPRS also provides an image processing facility for CBEL and other collaborating laboratories.

Research Projects

High Performance Biomedical Computing

R. L. Martino, Ph.D.
with C. A. Johnson, J. C. Pfeifer, E. T. Seidl, Ph.D., E. B. Suh, B. L. Trus, Ph.D., N. I. Weisenfeld, T. K. Yap, C. J. Lanczycki (DCRT/CBEL); J. I. Powell, J. D. Malley, Ph.D. (DCRT/DSB); B. R. Brooks, Ph.D., M. Hodoscek, Ph.D. (DCRT/MGSS); S. Erwin (Intel Supercomputer Systems Division); J. R. Caston, J. F. Conway, Ph.D., C. G. Hyde, Ph.D., A. C. Steven, Ph.D. (NIAMS/LSBR); A. Bax, Ph.D., M. G. Clore, Ph.D., F. Delaglio, G. Zhu, Ph.D. (NIDDK/LCB); B. K. Lee, Ph.D (NCI/LMB); D. Brewer, E. S. Gershon, M.D., Ph.D., L. R. Goldin, Ph.D. (NIMH/BPB); J. V. Haxby, Ph.D., J. Maisog, M.D. (NIMH/LPP); B. Horwitz, Ph.D., A. R. Macintosh, Ph.D., T. Zeffiro, M.D., Ph.D. (NIA/LN); R. E. Carson, Ph.D., M. E. Daube-Witherspoon, Ph.D., Y. C Yan (CC/NMD); J. van de Geijn, Ph.D., X. Huchen (NCI/DCT); A. Toga, Ph.D. (UCLA School of Medicine); A. T. Brunger, Ph.D., N. Carriero, Ph.D., P. Nadkarni, Ph.D. (Yale University); M. E. Colvin, Ph.D., C. L. Janssen, Ph.D. (Sandia National Laboratory); J. Ott, Ph.D. (Columbia University); J. Saltz, M.D., Ph.D.(University of Maryland); B. Narahari, Ph.D. (George Washington University); A. Choudhary (Syracuse University); O. Frieder, Ph.D. (George Mason University); N. Bauman, Ph.D., B. Venkataraghavan, Ph.D. (American Cyanamid Company); G. Weigand, Ph.D., S. L. Squires, Ph.D. (DARPA/CSTO)

The goals of the high performance biomedical computing program are to identify and solve those computational problems in biomedicine that can benefit from high performance hardware, modern software engineering principles, and efficient algorithms. This effort includes providing high performance parallel computer systems for the NIH staff and developing parallel algorithms for biomedical applications.

In addressing these computational challenges, CBEL is developing algorithms for a number of biomedical applications that can benefit from computational speedup including image processing of electron micrographs, protein and nucleic acid sequence analysis, nuclear magnetic resonance spectroscopy, x-ray crystallography, protein-folding prediction, quantum chemical methods, molecular dynamics simulations, human genetic linkage

analysis, medical imaging, and radiation treatment planning. The ultimate goal is to have high performance parallel computing facilitate the science that is done at NIH. While developing these computationally demanding applications, CBEL is investigating the following high performance computing issues: partitioning a problem into many parts that can be independently executed on different processors, designing algorithms so that delays of interprocessor communication can be kept to a small fraction of the computation time, designing the parts so that the computing load can be distributed evenly over the available processors or dynamically balanced, designing algorithms so that the number of processors is a parameter and the algorithms can be configured dynamically for the available machine, developing tools and environments for producing portable parallel programs and monitoring system performance, and proving that a parallel algorithm on a given machine meets its specifications.

As part of its high performance computing activity, CBEL operates the DCRT Highly Parallel Computer System. This Intel® Supercomputer Systems Division iPSC/860, obtained in collaboration with the DARPA Touchstone program, is a multiple instruction stream, multiple data stream (MIMD) distributed memory system. The system has 128 processor nodes with 16 megabytes of memory per node. A high-speed data pathway connects all the nodes of the system. Using a hypercube network topology, this hardware message routing facility connects each node as if it had a dedicated channel to all other nodes. Another important part of the system is the Intel® Concurrent I/O File System that provides a 10 gigabyte fast access mass storage facility for balancing disk input/output with the computational power of the processors. This consists of many small disks connected to I/O nodes that communicate with the processor nodes through the hypercube interconnect. Over the next 3 years, CBEL will be adding a next generation high performance parallel computer capable of providing 100 gigaflops of computing performance.

The President's Office of Science and Technology Policy (OSTP), through the Federal Coordinating Committee for Science, Engineering, and Technology (FCCSET), has initiated a multiagency High Performance Computing and Communications (HPCC) Initiative to strengthen the nation's research computing enterprise. Within the Department of Health and Human Services (DHHS), the focal point for the HPCC program is the National Institutes of Health. The CBEL high performance biomedical computing program is an important part of this national initiative.

The following sections describe the major collaborative high performance biomedical computing projects presently under development by CBEL.

Image Processing of Electron Micrographs

The method of high resolution cryoelectron microscopy in combination with three-dimensional computer image reconstruction allows the structure of herpesvirus and several other relatively large virus particles to be studied. However, even at moderate resolution, reconstructions of herpesvirus nucleocapsids pose a formidable computational challenge. An electron micrograph of a field of virus images can be treated as a set of 2D projections of the same particle with different orientations. If the orientations can be determined accurately, the 3D equivalent of the projection slice theorem can be applied to reconstruct the image.

The first computationally demanding step in the reconstruction, *FindView*, processes each particle separately, generating a list of possible orientations for that particle. The list specifies a set of possible results for the particle's rotational orientation in 3D space as well as the particle's translation within the projection. *FindView* uses the icosahedral symmetry of the virus to determine the radial lines in the projection's discrete Fourier transform (DFT) that represent the intersection of the projection plane with the equivalent views of the icosahedron. For every unique orientation, *FindView* computes the location of these "common lines," compares the value of the DFT along these lines, and selects the orientations which meet the least squares similarity criterion.

CBEL improved the performance of *FindView* during FY93 by implementing the origin correction code in parallel. In addition, a driver was written for submitting *FindView* jobs to the Parallel Batch Queuing System.

Following *FindView*, *Emicograd* refines the initial orientation estimates supplied by *FindView*, eventually producing a final set of particle orientations for the reconstruction. In practice, the refinement process actually requires many separate runs of *Emicograd*. Early runs build and refine the orientations for a basis set of particles, adding particles to the basis set incrementally. Once completely built, the basis set is then used to refine all other particles' orientations by running *Emicograd's* "local refinement" feature on each additional particle. The complete set of refined particles can then be put through a "global refinement" of all particles in the set. In order to compare the accuracy of the estimated origins of a given particle against the others in the set, *Emicograd* determines the mean-squared difference (residual) between the particles along their "cross-common lines," radial lines in the DFT of the projections representing where the symmetry-related views of the two particles would intersect. For each particle in the set, *Emicograd* iteratively determines the residual between that particle and the others in the set, finds the direction gradient of the minimum, and stops the iteration when further refinements yield less than 1% improvement in the average phase residual. *Emicograd* also iteratively determines the most self-consistent "handedness" polarity for the particles in the set, and automatically flips the particle images when necessary.

During FY93, CBEL staff, working in collaboration with the Laboratory of Structural Biology, NIAMS, completed the adaptation of *Emicograd* to the Intel iPSC/860 parallel computer. The global refinement code has been modified to distribute the images in order to process larger problems with more virus particles as shown in Figure 3. The global refinement scaled well on the parallel computer and achieved good speedup. In

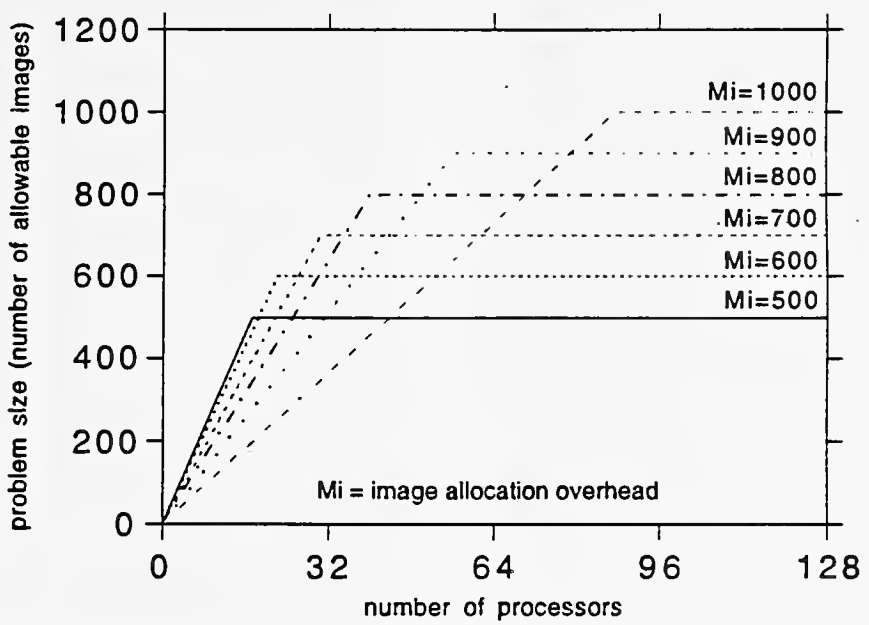


Figure 3. In the 3D reconstruction of icosahedral virus particles, orientation refinement is often the most computationally demanding task. The number of images which can be processed by the global refinement program is a function of the number of processors and the amount of image allocation overhead (a secondary array for the optimization problem). High performance computing makes possible reconstructions with larger numbers of virus particles, and consequently, techniques for image distribution must be considered.

response to the limitations of the original flipping methods (no more than 10 particles could be automatically flipped), CBEL developed, implemented, and tested a number of automatic flipping methods that overcome this barrier. The local refinement has been implemented on the iPSC/860 system as a separate program which distributes individual local refinement runs across the nodes. Recognizing that the true power of the parallel computer could not be realized without a means to conveniently construct large study queues and postprocess results, CBEL developed the sophisticated *Gradhost* front-end interface to the refinement engine.

During the coming year, CBEL plans to implement the final reconstruction stages in parallel. In their present form on the VAX™, these codes exist as four separate programs. The nature of the existing VAX™ implementation creates a limitation on the number of particles that can be reconstructed. This limitation has become a major obstacle as we seek to improve the resolution of the reconstructions by increasing the number of particles involved. Recent reconstructions have involved over 300 particles, and we view this increased-problem-size trend continuing as we seek to improve resolution and even reconstruct features which do not completely conform to the icosahedral model.

Protein and Nucleic Acid Sequence Analysis

As part of the human genome project, protein and nucleic acid sequence analysis focuses on identifying potentially important functional domains involved in gene regulation and chromosome organization. The identities of such sequences are elicited by multiple analytical approaches and require sequence comparisons between the analogous intergenic regions in multiple species and the recognition of unusual patterns of sequence within a single organism. When researchers discover new sequences, they are eager to search the database for sequences that are similar or relevant to their discoveries. In addition, researchers often search the database at regular intervals to keep up to date since new sequences are being added periodically.

The protein and nucleic acid sequences are maintained in a number of databases by different organizations. Some databases such as GenBank, SWISS-PROT, and EMBL (European Molecular Biology Laboratory) are well known internationally. These databases contain sequences from multiple species. For example, GenBank not only contains human DNA sequences but also plant, viral, bacterial, and other species sequences as well. Some sequence databases are derived from the above databases. One of these derived databases is AACC (amino acid class covering) Pattern Library which is maintained by Harvard University in Cambridge, Massachusetts. Each sequence in the AACC Pattern Library represents a family of sequences from the SWISS-PROT database.

In FY92, CBEL implemented a parallel version of a popular protein database search tool (PLSEARCH™) on its iPSC/860 hypercube computer. This tool was developed by Drs. Randall Smith and Temple Smith for searching, matching, and aligning newly generated protein sequences against the sequences in the AACC Pattern Library. To search the database, each query sequence must be compared against all the sequences in the database. In this application, each sequence comparison can be done independently. This is an important factor that was used in the design and implementation of the parallel version.

For the parallel version, we used the *manager-worker* approach to distribute and balance the computation load on each processor. In this approach, the *manager processor* initially distributes a query sequence and a sequence from the library to each *worker processor* where the similarity score is calculated. When each worker finishes the calculation, it sends the score back to the manager. Once the manager has received the score from a worker, it sends another pair of sequences to that worker for another calculation. This process continues until every query sequence has been paired and compared with all the library sequences. To minimize the time that the other workers have to wait for the last one to finish its comparison, the

library sequences are pre-sorted by their length in nonincreasing order before they are distributed. This is to ensure that the last sequence compared is the shortest one which requires the least amount of time. Using this approach, the load balancing is achieved since the workers are always loaded as long as there is work to be done.

In FY93, CBEL implemented another database search tool (GBSEARCH) using Gotoh's sequence comparison algorithm, which is based on the well-known Smith-Waterman's algorithm, for searching the entire GenBank for similarity sequences. A better approach is taken to implement this tool on the iPSC/860. Unlike the manager-worker, this new approach does not perform the load balancing for every query. It pre-determines the load for each processor based on the information of the current database. In this approach, the sequences in the original database are placed into one of the p = 128 *buckets* (smaller databases) so that the difference between the total length of the sequences in the smallest and largest buckets is minimized, where p is the largest number of processors in the hypercube. Once the original sequences have been placed in each bucket, each processor can search one or more buckets independently. This new approach is better than the previous one because it eliminates the communication time among the processors almost entirely. Communication is needed only at the very end to determine the global similarity sequences. In addition, it does not have the potential system bottleneck that is imposed on the manager.

Generally, only the best N records of sequence identifications and scores are saved for each search. These N records are kept in an inverted heap where the record having the lowest score is kept at the top. If the new record has a smaller score than the top one, it is discarded and the heap is not affected. However, when the new record has a higher score, it is used to replace the top one. In this case, the heap is adjusted. The amount of time that it takes to adjust the heap is in the order of $O(\log_2 N)$. As a result, the performance of both approaches also depends on the size of the heap (N). In the manager-worker approach, the manager could become a system bottleneck for a large N. On the other hand, in the bucket approach, it could take longer to determine the global N best records since each processor has more records to exchange between its neighbors.

To compare the performance of the manager-worker and the bucket approaches, the entire GenBank was used. The current release (rel. 76.0) has a total number of 111,911 sequences with 129,968,355 bases. In this release, the shortest sequence contains 1 base; the longest, 315,344; the average, 1,161; and the median, 449. For the query sequence, a median length sequence, T00361, was taken from the database. The serial search time is 1489.5 minutes. The computation timings of the two approaches are shown in the Table 2. These timings were based on a heap size of N=50.

In FY93, CBEL had also ported a multiple sequence alignment (MUSEQAL) program that was developed by M. P. Berger and P. J. Munson (DCRT, LSB) from the IBM® PC to its Intel® iPSC/860 computer and to the UNIX® workstation. MUSEQAL is a valuable and effective tool for analyzing evolutionary, functional, and structural relationships among protein sequences. This program randomly divides n pre-aligned sequences into two groups. Then, it aligns the sequences in one group against the other by freezing the alignment within each group. Thus, the alignment between the groups is optimized by using a two-dimensional Needleman-Wunsch type of algorithm. The resulting alignment, in turn, will be the starting point for the next alignment of a different pair of subgroups. Iteratively, an optimal overall alignment for all n sequences is thus approached.

As described by Berger and Munson, the multiple sequence alignment algorithm iteratively applies the pairwise sequence alignment type of algorithm. For each iteration, it randomly partitions the sequences into two groups. Then, it aligns the two groups against each other. Without restriction, there are $2^{(n-1)} - 1$ possible partitions to choose from, at each iteration, where n is the number of sequences. Each of these partitions can be aligned in parallel. This iterative approach was used in the design and implementation of the parallel version. In

Table 2. Timings for searching the entire GenBank (Rel. 76.0) for wEST01082 *Caenorhabditis elegans* cDNA clone CEESE52, which has 448 bases. See text for a description of terms used in the table.

	Time (min)		Speedup		Efficiency (%)	
Number of Nodes	*Manager Worker*	"Bucket"	*Manager Worker*	"Bucket"	*Manager Worker*	"Bucket"
2	1481	78	0	2	0	99
4	494	374	3	4	75	99
8	212	188	7	8	87	99
16	99	94	15	16	94	99
32	48	47	31	32	97	98
64	24	24	62	62	97	97
128	13	13	119	119	92	92

each iteration, each processor performs the alignment on a different partition. At the end of each iteration, the resulting global optimal alignment is used by all processors as the starting point for the next alignment.

These new versions (iPSC/860 and UNIX) allow the user to align more sequences than was possible on the PC since the parallel computer and the SUN workstation have more memory and greater computation power. In addition, CBEL had also developed a friendly graphical user interface for obtaining inputs from the user and for displaying the aligned sequences (Figure 4).

In FY94, CBEL will continue to maintain the parallel versions of PLSEARCH, PMUSEQAL, and GBSEARCH on its iPSC/860 hypercube computer. We will also investigate an efficient parallel algorithm for comparing and aligning two very long sequences.

Nuclear Magnetic Resonance Spectroscopy and x-ray Crystallography

This activity involves the development of parallel software tools for NMR spectroscopy and x-ray crystallography. This includes tools for the 3D structure determination and refinement of biomolecules using crystallographic data or NMR data.

In FY92, CBEL implemented the maximum entropy method (MEM) for 3D NMR data sets on the Intel iPSC/860 parallel computer. However, the program is not sensitive enough for the number of peaks present in the 3D data. There are many peaks of approximately the same intensity, and MEM is allowing an unacceptably large number of errors before there is a noticeable effect on the entropy function. In FY93, our collaborators investigated ways to overcome the sensitivity problems of MEM as well as alternatives to MEM. This effort will continue in FY94.

In the past year, CBEL continued to support Frank Delaglio of the NIDDK Laboratory of Chemical Biology in his development of a parallel Genetic Algorithm (GA) approach to spectral assignment, determining which signals in the NMR spectrum belong to which atoms in the protein. In his prototype spectral assignment scheme, he considered the assignment of the protein calmodulin, a 148-amino-acid sequence which had already been analyzed by manual methods. When run on 64 processors of the Intel computer, the GA method successfully identified the correct assignment in 12

hours time, evaluating over 8 million possible assignments in the process. The GA approach even discovered an error in the manual assignments.

CBEL also implemented a parallel version of the X-PLOR program system developed by Dr. Axel Brunger of Yale University and widely used by x-ray crystallographers and NMR spectroscopists throughout the world. In the coming year, work will continue on this large software development effort.

task of searching through all the possible conformation paths. CBEL is working to implement the *CHORUS* program on the Intel iPSC/860 parallel computer, so that a protein-folding simulation can be performed within a reasonable time. The computationally intensive part of this program is the calculation of the solvent accessible surface area of the protein as it progresses to its final structure (Figure 5).

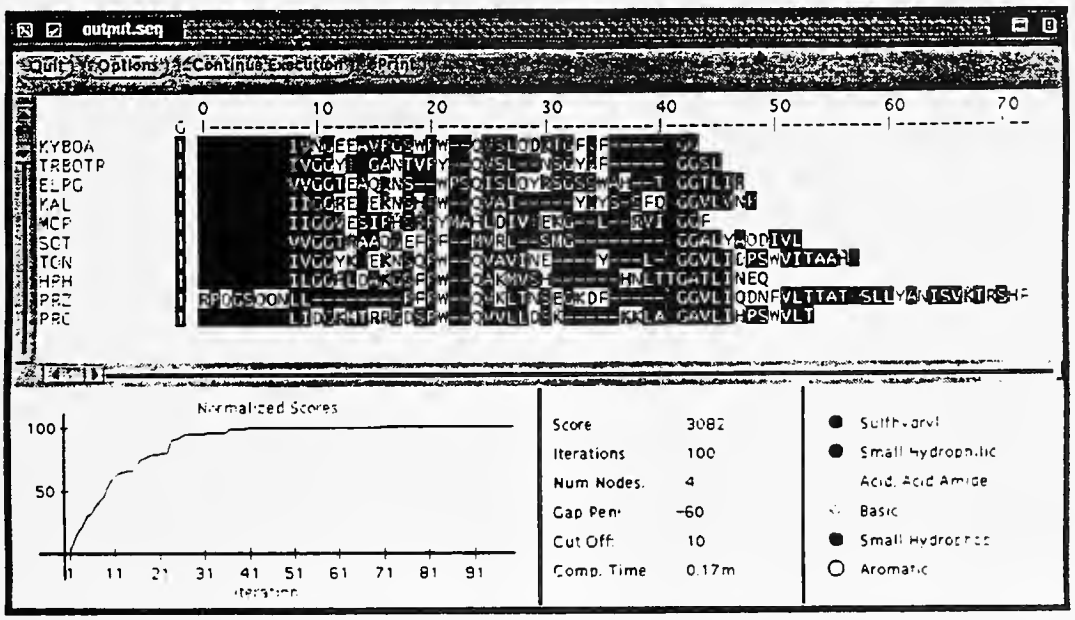


Figure 4. CBEL-developed graphical user interface for displaying multiple sequences aligned using the Berger-Munson MUSEQAL algorithm.

Protein-folding Prediction

The protein-folding problem is concerned with obtaining the 3D tertiary structure of a protein molecule when only its amino acid sequence is known. Dr. B. Lee of the Molecular Modeling Section, LMB, DCBDC, NCI, is developing a protein-folding simulation program called *CHORUS*. This program requires an enormous amount of computing time to simulate the folding of even a small-size protein molecule because of the complex

CBEL has implemented three surface area calculation algorithms on the Intel machine: Richmond's exact algorithm, Shrake and Rupley's approximation algorithm, and Lee and Richards' approximation algorithm. All three algorithms can be used in the *CHORUS* program to approximate the hydrophobic effects. Using 128 processor nodes to calculate the solvent accessible surface area of the lactate dehydrogenase protein molecule with the Shrake and Rupley's approximation algorithm, the Intel computer was 45 times faster than the IBM

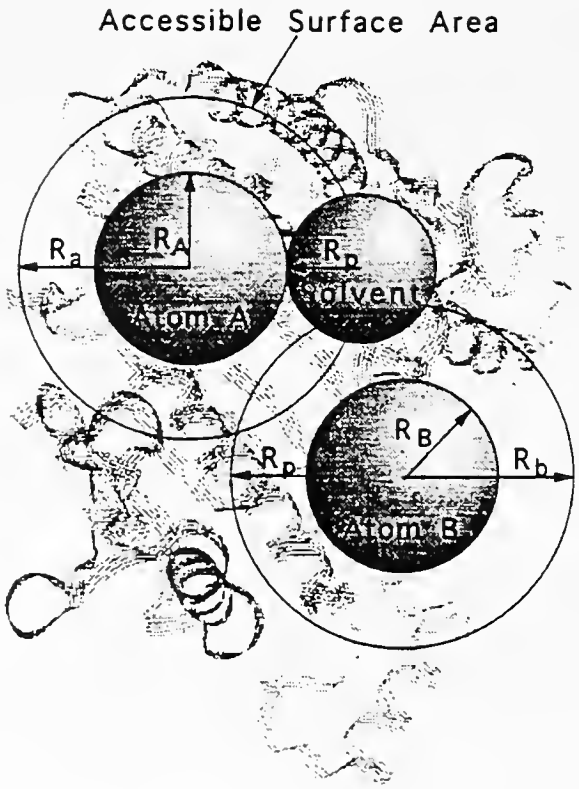


Figure 5. Calculating the solvent accessible surface area of a protein. Quantities R are van der Waals radii as shown. Solvent accessible areas are defined in terms of areas swept out by rolling a solvent "sphere" over spheres of protein atoms.

3090 300J and 105 times faster than the Convex C240. In FY93, CBEL implemented several new versions of static and dynamic load balancing methods for the Lee and Richard's algorithm. These include the spectral bisection, coordinate bisection, and quadtree algorithms. The effect of load balancing was to decrease the computation time by 11% when calculating the surface area of myoglobin. The improvement due to load balancing will increase with the size of the protein being simulated.

In FY94, CBEL will continue work on efficient parallel methods for calculating the solvent accessible surface area. In the coming years, CBEL will develop a parallel version of the *CHORUS* program with Dr. Lee.

Quantum Chemistry

The goal of this project is to develop *ab initio* quantum mechanical methods for use on massively parallel computer architectures. Unlike empirical force field (or molecular mechanics) or semi-empirical methods, *ab initio* methods are not parameterized, thus they may be used to describe previously unknown chemical systems with a high degree of accuracy. Unfortunately, however, *ab initio* methods are quite computationally expensive; thus, to date, they have been applied only to small chemical systems (usually fewer than 30 atoms). In order to treat systems of biological interest (greater than 100 atoms), computers with speeds at least in the gigaflop (billions of floating point operations per second) range will be needed. Computation time

increases according to N^4-N^6, where N is the number of atoms. Thus, a 3-fold increase in N may require a 100- to 1,000-fold increase in computer time. At present, massively parallel architectures provide the greatest hope of achieving the required speed economically.

Work to date has centered on implementing the Hartree-Fock Self-Consistent Field (SCF) approximation to the time-independent Schroedinger equation on multiple instruction stream, multiple data stream (MIMD) distributed memory parallel computers. In the SCF method, the molecular wave function is described by a determinant of single-electron functions, known as orbitals, which are themselves expanded in terms of a set of known functions (basis functions). The potential energy term arising from electron-electron repulsion is treated by calculating an effective field due to the average positions of the electrons. This field is varied in an iterative fashion until self-consistency is reached. The computational bottleneck in an SCF calculation is the construction of the Fock matrix, which depends on the calculation of $O(n^4)$ (where "n" is the number of basis functions) electron repulsion integrals (ERIs). In the traditional SCF approach, these ERIs are calculated once, written to disk, and then read back in every SCF iteration. This requires a great deal of disk space, however. An alternative method, the direct SCF method, obviates the need for large amounts of disk space by recalculating the needed ERIs every SCF iteration. Although the direct SCF method seems to be much more expensive than the traditional approach, it is possible to greatly reduce the number of ERIs to be evaluated each iteration. Further, since the information necessary to calculate all ERIs can be stored in memory on each node of a parallel computer, the direct SCF calculation may be easily parallelized.

The culmination of the first phase of a collaboration with Drs. Curtis Janssen and Michael Colvin of Sandia National Laboratories has been the development of a set of libraries for use in *ab initio* methods as well as a prototype program, *mpqc*, which makes use of these libraries. The programming language C was chosen for these libraries since it allows for a great amount of flexibility and portability. Current capabilities of *mpqc* include closed-shell and high-spin open-shell SCF energies and analytic first derivatives, Mulliken and Lowdin population analyses, and electrostatic potential determination. Molecular symmetry is used to reduce the cost of both the energy and gradient calculations. Minimum and transition state searches may be performed in both Cartesian and internal coordinates. Most important, however, is the ability to use distributed matrices, greatly increasing the size of calculation which can be performed. If a complete copy of each matrix had to be held on each node, the size of problem which could be treated would be determined by the amount of memory on each node, regardless of how many nodes were used. By distributing the matrices, the size of problem which can be treated is determined by the number of nodes. To our knowledge, *mpqc* is the only quantum chemistry package with this distributed matrix capability, and using this capability we have performed SCF calculations on systems with as many as 2,300 basis functions describing 125 atoms.

While single point SCF calculations are of some use, particularly in the determination of atomic point charges, what one most desires from *ab initio* methods are optimized structures. Only optimized molecular geometries are of use in the determination of most other molecular properties, as well as the energetics of chemical reactions. Given the great expense of each individual SCF calculation, it is imperative to optimize the geometry of a molecule in the fewest number of steps possible. To this end, very powerful optimization methods have been implemented in *mpqc*. Rather than being written in C, however, these methods have been developed in the object-oriented language C++. The usefulness of these methods can be demonstrated with one example from the scientific literature. The optimization of the molecule 7-thia-1, 3-diazabicyclo(3.3.0) octa-2,4-dione ($C_5H_6N_2O_2S$, Cambridge Structural Database designation ACTHCP) is a good benchmark of the effectiveness of an optimization method. Using the simplest set of

widely used molecular dynamics software packages allowing for better levels of theory (e.g., electronic polarization), longer simulations giving better statistics, and larger molecular systems.

Human Genetic Linkage Analysis

Human genetic linkage analysis focuses on mapping genetic loci, such as genetic markers and genes associated with inherited traits, to relative positions on the chromosome through statistical analysis of inheritance patterns in families (pedigrees). Knowing the location of genes and the corresponding genetic traits they produce allows researchers to discover patterns of the genomic organization with important functional consequences and to compare humans with other mammals. Detailed maps of the human genome should quickly lead to major human health benefits. For example, by identifying genes or regions of DNA involved in several diseases, including hereditary forms of cancer, Alzheimer's disease, manic-depressive illness, Huntington's disease, and cystic fibrosis, new methods of diagnosis and treatment can be developed. Equally important, the better understanding of human biology that would follow from these studies would contribute broadly to the treatment of most diseases.

One of the widely used computer programs for performing human genetic linkage analysis is LINKMAP. This program is able to infer the likely position of a disease gene by iteratively calculating its likelihood at a series of points along a map of a chromosome, relative to the position of known markers from a number of pedigrees. A figure based on probability theory, the *lod* score (log of the odds ratio), is calculated for each position indicating the statistical level of support for the specified map. A *lod* score of 3 or higher is a conventional value suggesting linkage. The position with the highest *lod* score is the most likely location of the gene. Although this program provides valuable data, its use is limited by its long computation time especially when large pedigrees or many markers are studied. Depending on the number of positions; the number, size, and complexity of the families; and the number of loci, analysis performed using LINKMAP often takes 20 hours to run on a SUN® workstation, and may take as long as a week. To obtain higher resolution genetic linkage maps, the number of calculated positions must be increased accordingly. This requires even longer computation time.

To shorten the computation time, CBEL had ported LINKMAP version 3.1 to the Intel® iPSC/860 parallel computer in FY92. The design and implementation of the parallel algorithm were based on the following observation. The calculation of the *lod* scores for all the suspected gene positions over all the families used can be summarized in the following pseudo-code. The data from each family are used to calculate a partial *lod* score for each position.

```
FOR map-position =1 to N DO {
        FOR family = 1 to M DO {
                calculate partial lod score
        }
}
```

These NxM partial *lod* scores represent independent tasks which can be performed by separate processors. To compute all the partial scores in parallel, the control decomposition technique is used. In this technique, one processor acts as a manager, distributing the necessary data to the worker processors to compute the partial *lod* scores. After the manager processor collects all the partial *lod* scores for each position from the workers, it combines them into a single *lod* score for that position. Using 128 processors on the Intel® iPSC/860, a LINKMAP computation that required almost four days on a SUN® workstation was reduced to less than an hour.

In FY93, CBEL upgraded its parallel version to match up with the sequential version 5.1 that it had acquired from the Baylor College of Medicine in Houston, Texas. This particular sequential version has a major algorithmic modification that will provide better performance. As a result, the new parallel version also yields a better performance.

position of a disease gene by iteratively calculating its likelihood at a series of points along a map of a chromosome, relative to the position of known markers from a number of pedigrees. A figure based on probability theory, the *lod* score (log of the odds ratio), is calculated for each position indicating the statistical level of support for the specified map. A *lod* score of 3 or higher is a conventional value suggesting linkage. The position with the highest *lod* score is the most likely location of the gene. Although this program provides valuable data, its use is limited by its long computation time especially when large pedigrees or many markers are studied. Depending on the number of positions; the number, size, and complexity of the families; and the number of loci, analysis performed using LINKMAP often takes 20 hours to run on a SUN® workstation, and may take as long as a week. To obtain higher resolution genetic linkage maps, the number of calculated positions must be increased accordingly. This requires even longer computation time.

To shorten the computation time, CBEL had ported LINKMAP version 3.1 to the Intel® iPSC/860 parallel computer in FY92. The design and implementation of the parallel algorithm were based on the following observation. The calculation of the *lod* scores for all the suspected gene positions over all the families used can be summarized in the following pseudo-code. The data from each family are used to calculate a partial *lod* score for each position.

```
FOR map-position =1 to N DO {
        FOR family = 1 to M DO {
                calculate partial lod score
        }
}
```

These NxM partial *lod* scores represent independent tasks which can be performed by separate processors. To compute all the partial scores in parallel, the control decomposition technique is used. In this technique, one processor acts as a manager, distributing the necessary data to the worker processors to compute the partial *lod* scores. After the manager processor collects all the partial *lod* scores for each position from the workers, it combines them into a single *lod* score for that position. Using 128 processors on the Intel® iPSC/860, a LINKMAP computation that required almost four days on a SUN® workstation was reduced to less than an hour.

In FY93, CBEL upgraded its parallel version to match up with the sequential version 5.1 that it had acquired from the Baylor College of Medicine in Houston, Texas. This particular sequential version has a major algorithmic modification that will provide better performance. As a result, the new parallel version also yields a better performance.

For FY94, CBEL will continue to maintain this program on its parallel computer and provide support to users throughout NIH.

Functional Neurological Image Analysis

CBEL is investigating the use of high performance computing technology and image processing techniques to solve problems in functional neurological image processing. This work began in FY92 in collaboration with researchers from NINDS and NIA and has continued in FY93 with researchers from NIMH and the UCLA School of Medicine. Current work involves many aspects of functional neurological image processing using Positron Emission Tomography (PET) as the primary modality to measure and analyze regional cerebral blood flow (rCBF) activation.

One of the biggest problems facing researchers in this area is image registration. In order to effect automated processing and analysis of PET scans, it is necessary to map those images into the coordinate system of a brain atlas. Once registered to an atlas, the PET activation data can be correlated with the functional areas of the brain in which these activations have occurred. While the registration of two PET images from the same subject can be performed with a linear, rigid transformation, the registration to a standard stereotactic brain atlas of PET images from brains of different shapes and sizes

can only be achieved through a class of nonlinear deformations collectively called "warping."

In FY93, CBEL began investigating warping techniques for the registration of three dimensional PET images. Along with Dr. James Haxby of NIMH, CBEL initiated a collaboration with Dr. Arthur Toga of the UCLA School of Medicine to implement and refine his unique warping method. Since accurate, reproducible 3D registration can pose quite a computational challenge, CBEL has begun implementing Dr. Toga's method on its 128-node Intel® iPSC/860 highly parallel supercomputer. In FY94, CBEL plans to continue development of this software as well as investigation of new techniques for 3D image registration.

NIH scientists are currently investigating the functional connectivity of regions in the human brain using PET images. Two brain regions which are functionally associated will show a highly correlated level of regional cerebral blood blow activity, as measured by PET in a given group of subjects under specific experimental conditions. Principal component analysis (PCA) produces orthogonal maps of these highly correlated regions by generating the eigenfunctions of the ensemble of brain images. In FY93, CBEL implemented the PCA algorithm using the "duality" approach, one which greatly reduces the computational and storage requirements.

In addition to new research and software development, CBEL has continued support of its high performance versions of portions of the Statistical Parametric Mapping (SPM) software. SPM was originally developed by Dr. Karl Friston at the Hammersmith Hospital in London and written in the script language of the MATLAB system. In FY92, CBEL members rewrote two key parts of SPM, the plastic normalization and ANCOVA, in the C programming language. Through this translation, an enormous computational speedup was achieved. In FY93, CBEL continued to support new and current users of its SPM software and has worked to make the software available to other scientists at NIH. Part of this effort was the enhancement of existing input/output routines to allow full interoperability and interuse of code and data between the DEC® Ultrix™ and SUN® Microsystems SUNOS platforms. This enhancement brings the SPM software and the results of its computations to researchers using both types of systems. CBEL plans to continue its support for SPM in FY94.

Reconstruction of Positron Emission Tomography Images

Positron Emission Tomography (PET) is the most promising tool for biochemical imaging today. It can provide diagnostic information that x-ray computed tomography (CT), digital subtraction angiography (DSA), ultrasonography, and magnetic resonance imaging (MRI) cannot. PET can provide clinicians with chemical and metabolic information and define patterns of chemical change in the disease under study.

A PET image is formed through a computational reconstruction process. The quality of the resulting image and the computation time needed to produce it depend on the chosen reconstruction algorithm. Traditionally, Fourier methods (e.g., the filtered backprojection algorithm, which is fast but can lead to artifacts) have been used. Another class of methods known as algebraic methods, an example being the expectation maximization (EM) or maximum likelihood (ML) algorithm, are known in theory to yield more accurate reconstructions or equivalent reconstructions with lower patient dose. The algebraic methods have not been used in the past because of the long computation time and the large amount of memory required to implement them.

With the availability of high performance parallel computer technology, NIH scientists can now consider applying the ML algorithm to the problem of fully 3D reconstruction. The new generation of PET scanners allows for the retraction of the lead septa shields, which prevent coincidence events from being detected outside the axial plane of the emission. Retracting the septa increases the angle over which coincidence events are accepted, and consequently improves the detector sensitivity and the count rate. However, the amount of detected scatter and random events also increases with wider

acceptance angles, and a current debate focuses on whether retracted septa scanners can lead to improved reconstruction quality. Another drawback to retracting the septa is that the size of the reconstruction problem grows enormously, especially with algebraic approaches. In a 3D ML reconstruction using typical scanner geometries, the number of projections (rays of coincidence events) grows by an order of magnitude, and the size of the probability matrix, which is used throughout the ML reconstruction, can grow by four orders of magnitude or more.

During FY93, NIH scientists applied the Bergstrom method of scatter correction to the 2D ML reconstruction code, which has been implemented on the Intel® iPSC/860. Since the scatter fraction for brain sized objects increases from 10-15% using current-generation scanners to as much as 50-60% with the septa removed, a scatter correction must be performed in order to achieve accurate images from a full 3D reconstruction. In FY94 we hope to implement some form of the 3D reconstruction algorithm on the parallel computer. Although NIH does not yet own a new-generation scanner with retractable septa, the computational geometry of the 3D problem on the parallel computer can be formulated based on the published performance figures of new-generation scanners already on the market. This problem presents a number of computational challenges to the parallel developer. For example, the probability matrix is so large that it cannot be stored, even if compacted. After we solve the basic parallel problem, we plan to compensate for attenuation, normalization, randoms, and scatter in the 3D ML model.

Radiation Treatment Planning

In radiation treatment planning, a radiation oncologist tries to determine the optimum placement, blocking and intensity of beams such that the body volume to be irradiated gets the maximum dose while minimizing damage to surrounding tissue. Computers can be used to greatly improve the success of this task. A series of images, such as those obtained from a CT scanner, are read into the computer and different volumes are then identified. These would typically include bone, lung, internal organs, spinal cord, and the tumor. A radiation beam placement plan is then specified and a simulation is performed by the computer. The result of this simulation is a series of 2D contour maps that shows the percent of maximum dose for each area of the body. The radiation oncologist uses these contour maps to determine the most effective beam plan.

MacTPS is a Macintosh®-based radiation treatment planning system developed by Dr. Jan van de Geijn and Huchen Xie of NCI. The Macintosh® provides a good graphical interface for preparing the images and specifying the beam plan but the simulation of the treatment takes an unacceptably long time. CBEL plans to implement the time-consuming simulation part of this program on the Intel® iPSC/860 parallel computer. In FY93, we set up communication between the parallel computer and the Macintosh® computers via TCP/IP internet sockets. In FY94, CBEL will determine the best method to implement a parallel version of the computationally intensive parts of MacTPS along with making the necessary changes to MacTPS to interface to the Intel® system.

Chemical Structure-Activity Relationship Studies

The search for new drugs in the pharmaceutical industry is a costly and time-consuming process. To reduce cost and shorten the length of time, it is important to identify new chemical compounds that are worthy for clinical evaluation as early as possible in the design and development process. As the chemical compounds advance from one step to the next in this process, the cost of evaluating them is higher and the evaluating time is longer. The early steps of the process can be broadly classified into three classes: (1) computer screening, (2) *in vitro* screening, and (3) *in vivo* screening.

Traditionally, before the computer screening techniques became available and the number of known chemical compounds was small, the *in vitro* screening technique was used as the first step to screen these compounds for biological activity.

When the number of compounds became large, it was not practical to screen all the compounds as it was too costly and time-consuming. As a result, only a small number of compounds was randomly selected for screening. Using this method, the rate of discovery was generally unsatisfactory.

To improve the rate of discovery, some computer screening techniques were developed. These techniques use mathematical methods to screen the chemical compounds in the database for a particular activity. This approach generally involves the use of biological test data from previous studies and chemical structure data to create a system for predicting the biological activity of a new compound. The compounds selected using this method are further tested with *in vitro* screening, where tests are carried out in the laboratory. The remaining compounds are further tested with *in vivo* screening where the tests are carried out on animals.

In collaboration with a researcher in the pharmaceutical industry, CBEL is implementing a computer screening method that provides a higher rate of discovery than the random screening technique on its iPSC/860 hypercube computer. The combination of this method and the power of parallel machine should shorten the screening time and reduce the cost of *in vitro* and *in vivo* screenings enormously. This reduction of time and cost becomes very significant when a large database of chemical compounds has to be screened routinely for different biological activities.

The basic idea that we use to design the activity prediction system is based on the assumption that chemical compounds with similar structures would also have similar activities as well. There are many ways to represent the chemical structures. We use a simple but effective way in which a structure is represented by a vector of atom-pair descriptors. Each descriptor represents the distance and the properties of a pair of atoms. The property of each atom, in turn, contains information such as the atom type, the number of bonding pi-electrons, and the number of nonhydrogen neighbors. To predict the activity level of a chemical compound in the database, we compare its structure with all the structures of the compounds in a small training set that have known activity levels. The compound that has the highest similarity score is the most likely to have the highest level of activity that contains in the training set.

Based on a small set of training compounds, this technique was able to predict with 73% accuracy. The test training set contained 121 compounds and 42 of them were active (containing varying level of activity). The preliminary performance of the parallel implementation showed very encouraging results.

In FY94, we will evaluate this technique using a larger chemical structure database of our collaborator. Based on the new results, we will work on refining this technique further.

Remote File Access and Communication System

CBEL continues to investigate ways to bring the Intel® iPSC/860 parallel computer into widespread use in the NIH distributed computing environment. Computations performed on the parallel computer generally require fast input/output (I/O) transfers as they involve large amounts of data as input and generate large amounts of output data. While the iPSC/860 has its own high-speed disk system, the user's data and programs are, more often than not, stored on either the user's own workstation or on a central file server. The system software developed by Intel® and supplied with the iPSC/860 provides a convenient, transparent interface for programs running on the hypercube that need to read files from a user's workstation. Unfortunately, this system requires that all data be relayed through the parallel computer's front-end processor, the System Resource Manager (SRM), instead of traveling directly between the hypercube and the remote host workstation. The SRM is greatly overburdened since, in addition to its other functions, it is also responsible for performing this task for every process running on the hypercube. The result is an I/O bottleneck that is often unacceptable.

CBEL has developed a set of symmetric routines that allow processes running on the

iPSC/860 to access workstation files and allow processes running on the workstation to access iPSC/860 files. As was expected, the file I/O operations were significantly faster using CBEL's software, the Remote File Access and Communication System (RCOMM). During FY93, CBEL used this library of routines to provide fast, robust file I/O to applications that are I/O intensive. In addition, the library was extended with a feature that allows the custom tailoring of its internal buffer sizes. This reduces communication overhead and allows the programmer to adapt the system to the size of the problem at hand. CBEL plans to continue to support, maintain, and extend RCOMM during FY94.

The Parallel Batch Queuing System

To make most efficient use of any computer system, it is necessary to keep as many jobs as possible running on it at all times. It is also desirable to allocate system resources in such a way that all users have equal opportunity to get work done. The easiest way to satisfy these two (sometimes conflicting) requirements is to have some form of queuing system allocate resources to all users. The Intel® iPSC/860 system has as part of its operating system the Network Queuing System, or NQS. NQS suffers from several deficiencies, however. First, jobs can only be submitted from and run on the System Resource Manager (SRM), which is a 386-based PC. NQS (as presently implemented) makes no use Intel's remote host software which allows jobs to be run from UNIX® workstations connected to the SRM via the ethernet. A second major deficiency is that jobs are removed from NQS's queues in a first-in, first-out manner; thus, if a job at the top of the queue is too large to be run, no smaller jobs behind it will be run, wasting available resources. Finally, NQS is of no use in controlling interactive use of the computer. For these reasons it was decided that a new queuing system was needed.

The Parallel Batch Queuing System (PBQS) developed by CBEL is based on the Multiple Host Batch Queuing System written by Curtis Janssen at the University of Georgia. PBQS corrects all of the deficiencies of NQS. Users may submit jobs from any UNIX® workstation which supports the Remote Procedure Call (RPC) protocol, and jobs may be run on any workstation for which Intel's remote host software is supported. PBQS also uses a more sophisticated algorithm for deciding which job to run, so more efficient use of the system is possible than when NQS is used. Finally, PBQS can be used to perform several useful administrative tasks, including limiting the size of nodes which users can allocate at different times of the day, reserving the entire computer for the use of one user, and keeping accounting records for each user.

While PBQS is remarkably stable, development continues in an effort to make it more user friendly. A graphical user interface has been developed which allows users to more easily monitor the progress of their jobs. Also, the scheduling algorithm continues to evolve in such a way that light users of the system achieve minimum turn around time, while keeping system usage at a maximum level. Finally, a library interface to PBQS has been developed which allows users to access PBQS functions from software they have written.

Structural Biology: Image Processing of Electron Micrographs

B. L. Trus, Ph.D.
with A. C. Steven, Ph.D., E. Kocsis, Ph.D., F. Booy, Ph.D., J. Conway, Ph.D., M. Misra, Ph.D., A. Makhov, Ph.D., M. Cerritelli, Ph.D., J. Caston, Ph.D. (LSB/NIAMS); J. Brown, Ph.D., W. Newcomb (University of Virginia)

This project uses image processing techniques to analyze electron micrographs. To answer important questions in structural biology, it is necessary to obtain relatively high resolution 2- and 3D structural information about biological macromolecules. While atomic or near-atomic resolution information traditionally has been available by x-ray crystallography for some small molecules and proteins, the overwhelming majority

of biological macromolecules are not crystalline, or are too large and therefore not amenable to 3D crystallography.

Biological specimens can, on the other hand, be visualized in the electron microscope using a number of specimen preparation techniques. Negative staining and shadowing, which both use heavy metals, are two traditional approaches to increasing contrast to show the biological macromolecule's structure. Cryoelectron microscopy, a newer technique, attempts to preserve "native" structure by surrounding the specimen with a layer of ice. Collaborative studies with LSB, NIAMS are currently under way on a number of projects, whereby the electron micrograph images are computationally corrected, combined, averaged, reconstructed, or in some way computationally enhanced to improve the signal-to-noise ratio or to increase the interpretability of the structures being visualized. "Cryo" images are typically lower contrast and require greater computer processing to achieve satisfactory results.

Sometimes the image processing results can be combined with amino acid sequence analysis to yield additional information about the macromolecular structure. Sequence analysis uses the one-dimensional amino acid sequence of proteins together with both Fourier analysis and other predictive algorithms to attempt to identify parts of the sequence that may have a regular structure and to predict 3D relationships.

Of particular interest to our research is the understanding of viral structures. At present we are continuing our efforts to investigate the structure of a large animal virus, human herpes simplex virus type 1. We are completing the localization of the major capsid proteins. Using the 3D icosahedral reconstruction technique, we apply the symmetry of these virus particles to both find the orientation of randomly oriented capsid particles (in ice) and combine many particles into a 3D reconstruction. Biological material for these herpesvirus reconstructions is provided through a collaboration with researchers at the University of Virginia, Charlottesville. The electron microscopy is performed in LSB, NIAMS. Interpretation of our 3D reconstructions is performed jointly by all collaborators.

Starting with the precursor herpes capsid (B-capsids), we have studied specimens with three different monoclonal antibodies. In addition, we have studied degradation products (e.g., guanidinium HCl or urea treatment) with the goal of determining the 3D location of the seven major capsid proteins. Difference 3D reconstructions, for example, clearly show that one protein, VP26, is bound on the outer tips of the hexons.

Future work on this project involves the use of additional antibodies to confirm our localization experiments of other major proteins, and an attempt to increase the resolution of our results substantially. The computational demands of the 3D reconstructions have prompted the use of DCRT's iPSC/860. This year, progress has been made in the use of a new *Gradhost* program to perform orientation searching and global refinement of orientations (see the section on High Performance Biomedical Computing).

Five other collaborative projects in structural biology are currently in progress. We are using similar 3D reconstruction techniques to study the structure of icosahedral bacteriophage T7 structure and of L-A virus (from yeast). Another project involving the 3D reconstruction of the cell wall of *Bordetella pertussis* has been completed. Two 2D projects involve a study of the connector proteins of T7, and Filamentous Hemagglutinin (FHA) from *Bordetella pertussis*. In the later study, amino acid sequence data may be combined with image processing results to yield additional useful information about the macromolecular structure.

Biomedical Image Processing

B. L. Trus, Ph.D.
with M. Vivino (DCRT/CBEL); K. Kempner, D. Adams, Ph.D. (DCRT/DSB); M. Datiles, M.D., A. Mahurkar, L. Lopez, M.D., B. Magno, M.D., S. Lassa, M.D., (NEI/OGCS); M. Jones, M.D., E. Tucker (NHLBI)

This project uses sophisticated image processing techniques to analyze biomedical images.

The goal is to establish collaborations with biomedical experts who require new algorithms and possibly new hardware capability to solve difficult imaging problems. Typically, complex new mathematical algorithms as well as new combinations of existing algorithms are utilized. We attempt to integrate the best computer platform for each problem with the desired goal of the project, using such diverse computers as an Apple® Macintosh®, a DEC® VAX™ or Alpha, a SUN® workstation, or an Intel® iPSC/860 supercomputer.

quantitate lens opacities (cataracts). In one system the use of images produced by the Scheimpflug principle evaluates nuclear opacities. The system developed in two directions. The first aspect developed was for the online capture of images. The second aspect of the system was image analysis. This includes routines to automatically locate three anatomical regions of the lens. Software then computes density measurements within the regions. During this past year NEI and DCRT presented a poster at the Advanced Research in Vision and

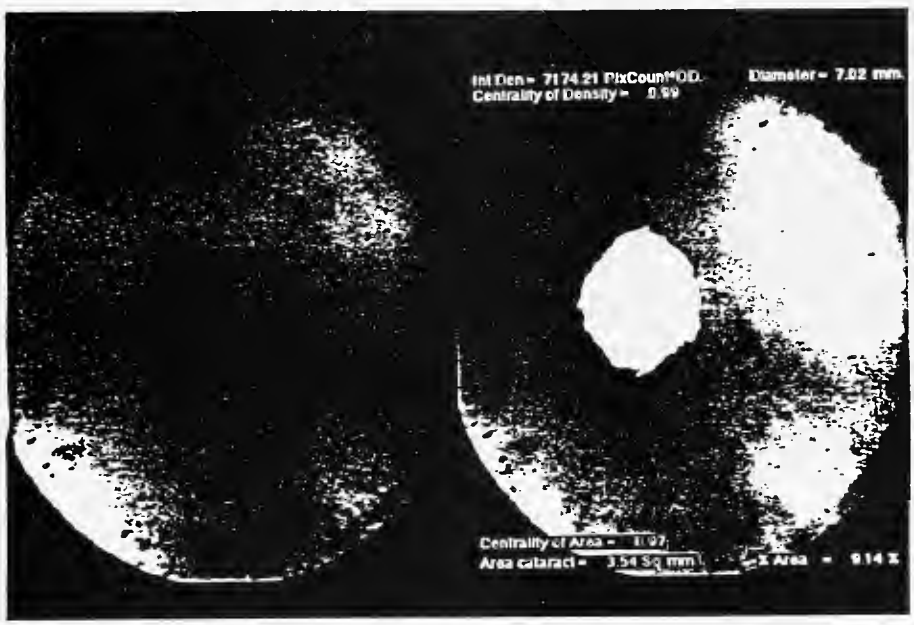


Figure 6. Retroillumination image before and after analysis. These images are frontal plane views of a lens with a cataract. The right side shows areas considered as cataractous and the quantitative analysis of area, density, and centrality.

Two current projects include ophthalmic image acquisition and analysis and ultrasound image analysis. In the first project, research has been ongoing between the National Eye Institute (NEI) and DCRT in the area of computerizing instrumentation and the automatic analysis of anterior eye segment images. The goal in developing computer-based NEI instrumentation is two fold: (a) to provide accurate and reproducible numerical information, and (b) to develop image analysis in a user-friendly and systematic way.

We are developing several systems to

Ophthalmology conference. The poster presented the first year's clinical data produced from the instrument. Results from these tests were considered quite significant in that the instrument is sensitive enough to show cataract progression in 1 year. No commercially marketed instrument has shown this sensitivity.

In addition to the Scheimpflug optics, NEI observes cataracts with the retroillumination system, which is more useful for posterior or anterior opacities. We developed a software-based system for analysis of retroillumination images (Figure 6).

These two devices can now be used to observe the effects of anticataract drugs or used for pathological grading.

We have also offered support for software being developed that assists in the evaluation of the corneal endothelial cells taken with a specular

actually a simulation of flow velocity. The goal for this project is to use an echo Doppler color mapping system and image the same artery or organ from multiple locations and orientations, then to use 3D calculations and reconstruct a true flow profile. The method being developed should allow not only better

Figure 7. Specular microscope view of corneal cells. Pathological cells are noted by nonuniform shapes, such as the lack of a characteristic hexagon.

microscope (Figure 7). This is done by a semi-automated technique that performs shape analysis on tracings of specular microscope images. Video capture of specular images is under consideration but is complicated by the low light level conditions associated with this instrumentation.

A second major collaborative effort with NHLBI as well as with DCRT/DSB is the measuring of blood flow velocity in arteries, and possibly through heart valves, noninvasively. Current ultrasound technology allows physicians to view flow approximately, but not quantitatively. Present systems provide a color display of flow which is

calculation of velocity profiles, flow volume and resistances, but also estimations of pressures across valve orifices and stenotic arteries and for other purposes.

We have procured necessary equipment and have constructed a phantom to test our algorithms. We have succeeded in transferring various flow velocity images, produced by the HP® SONOS® ultrasound system, into separate digital images of structure and flow velocity in our computer. We have also succeeded in obtaining data from a 3D position/orientation measurement system. We hope our approach will be useful to manufacturers and

users of echo ultrasound imaging systems.

Future projects include collaborating in the development of computer systems to analyze light microscopy images (including performing real time 3D reconstructions), as well as analysis of images from PET, SPECT, and MRI. We anticipate requests for collaboration in other "high tech" biomedical imaging projects, and we will participate to the extent that resources permit.

Publications and Presentations

Conway J. F., Trus B. L., Booy F. P., Newcomb W. W., Brown J. C., Steven A. C. Effects of radiation damage on frozen hydrated capsids of HSV-1, In Bailey G. W., Bentley J., Small J. A., eds. Proceedings of the 50th Annual Meeting of the Electron Microscopy Society of America, Electron Microscopy Society of America, San Francisco 1992; 532-3.

Havlin S., Kiefer J. E., Trus B., Weiss G. H., Nossal R. Numerical method for studying the detectability of inclusions hidden in optically turbid tissue, *App Opt* 1992; 32:617-27.

Kocsis E., Trus B. L., Steven A. C., Smith P.R., Hannah J. H., Brennan M. J., Kessel M. Orientation of porin channels in the outer membrane of *Bordetella pertussis*, *Mol Micro* 1993 (in press).

Makhov A. M., Trus B. L., Conway J. F., Simon M. N., Zurabishvili T. G., Mesyanzhinov V. V., Steven A. C. The short tail-fiber of Bacteriophage T4: molecular structure and a mechanism for its conformational transition, *Virology* 1993; 194:117-27.

Martino R. L. The NIH High Performance Computing and Communications Program. Presented at the Supercomputing '92 Conference, Minneapolis, Minnesota, November 1992.

Martino R. L. Parallel Computing in Structural Biology Research. Presented at the Drug Information Association Workshop on Research Perspectives in Structural Biology and Chemistry, Orlando, Florida, January 1993.

Martino R. L. High Performance Computing in Structural Biology and Medical Imaging. Presented at the Workshop on Grand Challenge Applications and Software Technology, Pittsburgh, Pennsylvania, May 1993.

Misra M., Conway J. F., Trus B. L., Steven A. C. Determination of nucleic acid content of viruses by electron spectroscopic imaging and correlation averaging. In: Bailey G. W., Rieder C. L., eds. Proceedings of the 51st Annual Meeting of the Microscopy Society of America, Microscopy Society of America, San Francisco 1993 (in press).

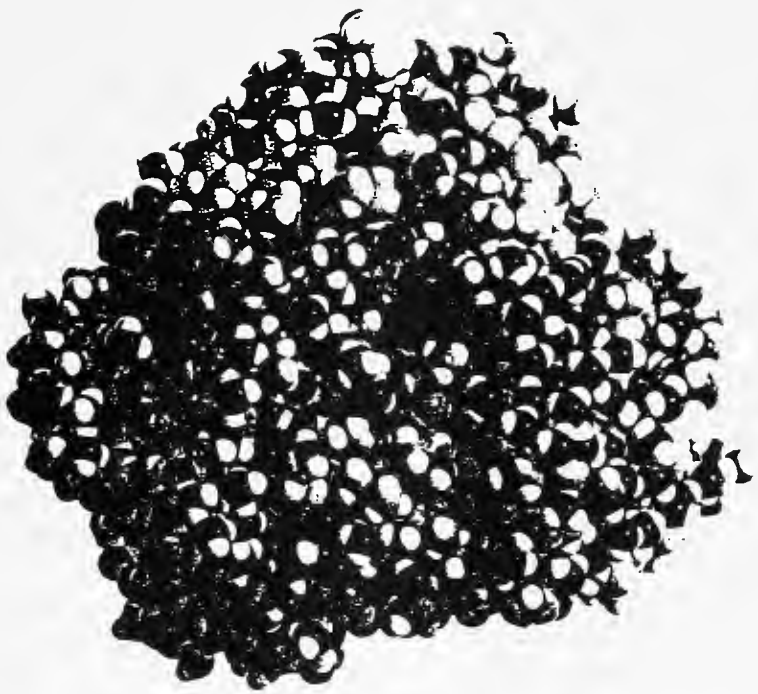
Misra M., Conway J. F., Trus B. L., Steven A. C. Determination of nucleic acid content of viruses by electron spectroscopic imaging and correlation averaging. In: Bailey G. W., Rieder C. L., eds. Proceedings of the 51th Annual Meeting of the Electron Microscopy Society of America, Electron Microscopy Society of America, San Francisco 1993; 570-1.

Newcomb W. W., Trus B. L., Booy F. P., Steven A. C., Wall J. S., Brown J. C. Structure of the herpes simplex virus capsid: molecular composition of the pentons and the triplexes, *J Mol Biol* 1993; 232:499-511.

Suh E. B., Lee B., Narahari B., Choudhary A., Martino R. L. Parallel computation of solvent accessible surface area of protein molecules. In: Proceedings of the Seventh International Parallel Processing Symposium, IEEE Computer Society, Washington, D.C. 1993; 685-9.

LSB

Laboratory of Structural Biology



Laboratory of Structural Biology

V. Adrian Parsegian, Ph.D., Acting Chief

Motivated by the need to merge different kinds of structural studies already being conducted in the division, the Laboratory of Structural Biology (LSB) was created this past year out of four diverse groups. A *Section on Molecular Forces*, led by Dr. V. Adrian Parsegian, is primarily concerned with the direct measurement of forces between and within macromolecular and cellular structures. From knowing how molecules interact, one expects finally to understand the recognition and specificity essential to cell function; one can begin to design molecules for specific effects and functions. This group shares laboratory space and research partners with NIDDK to forge a strong connection between theoretical ideas and laboratory practice. The effort of Richard Feldmann builds on his pioneering experience and success in molecular graphics and uses a wide variety of computing platforms to attack questions of protein folding and organization. The *Analytical Biostatistics Section*, headed by Dr. Peter Munson, examines the rapidly growing genome-sequence and protein-structure databases to test whether sequence/structure correlations can be strengthened to become a basis for structure prediction. By testing the limits of statistical methods, and recognizing the multiple variables that go into correlation, one expects to optimize the search for molecular information buried in massive amounts of data.

The *Molecular Graphics and Simulation Section* headed by Dr. Bernard Brooks, author of the widely used CHARMM molecular dynamics program, examines molecular motion and interaction through combined strategies of molecular mechanics and quantum mechanics. Beginning with known or postulated structures, one observes changes in molecular organization with molecular contact, with exposure to water, or with time-dependent interaction with substrates.

Members of this laboratory collaborate extensively with intramural and extramural scientists. A DCRT course "Physical Forces Organizing Biomolecules" attracted over 100 NIH researchers last spring, while a 1-day version in conjunction with the Biophysical Society meeting filled Lipsett Amphitheater. The force measurement techniques that have created this subject are now rapidly being adopted by several laboratories around the world. A meeting at the NIH on High Performance Computing in Chemistry, jointly sponsored by DCRT and the Pacific Northwest Laboratories, enjoyed international attention. The molecular dynamics algorithms, linked now to massively parallel high-speed computers, are enjoying use in a variety of practical and fundamental studies.

The LIGAND program, developed in the Analytical Biostatistics Section, continues to enjoy wide usage, especially now with its adaptation to Macintosh® computers.

Members of the laboratory have taught several specialized courses and have actively participated in the meetings of the NIH Structural Biology interest group.

Section on Molecular Forces

V. Adrian Parsegian, Ph.D., Head

Direct Measurement of Forces Between Membranes or Macromolecules

Use of the osmotic stress method, developed by our group to measure directly forces between membranes or between macromolecules, has spread rapidly this past year to several laboratories in Europe and in North America. One result of this recent proliferation in practice has been to advance the idea that as molecules or membranes approach contact, the important work of approach involves removal of organized water solvent from the apposing surfaces. These "hydration forces" are increasingly recognized to act in materials as diverse as lipid bilayers, proteins, DNA double helices, and stiff polysaccharides.

The growing catalog of information about interactions continues to create a new logic for thinking about molecular recognition and folding. It should eventually provide needed information to be used in the design of drugs targeted to specific sites by the strength of specific intermolecular forces.

During the current year we have concentrated on forces between proteins – specifically examining native and reconstituted collagen fibers at various temperatures, pH, ionic conditions, and in the presence of several small solutes. Unexpectedly, salt does not fully penetrate into the space between collagen triple helices. Consequently, we now realize that osmotic pressure applied from outside by the excluded salt is an important component of collagen fiber assembly and stability. Force measurements have demonstrated that temperature-favored assembly of the fibers is driven by water-mediated hydrogen bonding between the apposing polar residues. This is in contrast to "the hydrophobic effect," usually invoked to explain assembly of proteins. One can now think of a competition between repulsive and attractive hydration forces, depending on how well protein surfaces match each other.

A new kind of sensitivity of hydration forces between DNA molecules on small solutes has been discovered. We have explained an unusual "re-entrant" phase transition in lipid bilayers which go from liquid to gel and again to liquid form. The osmotic stress technique has been extended to measure forces between spherical particles in order to create convenient models for the "molecular crowding" phenomenon important in the regime of the intracellular milieu.

Forces Between DNA Double Helices

Methods

Standard DNA and collagen purification and characterization methods have been used in sample preparation. A new membrane filtration method has been developed to form highly ordered films of reconstituted collagen. To measure the forces between macromolecules or between membranes, the osmotic stress technique has been used to bring the molecules or membranes together into an ordered structure. X-ray diffraction has been used to measure molecular separation as a function of the applied osmotic stress and temperature.

Impact/Value

We now have a significant set of direct force measurements between biological macromolecules. This progress in understanding of underlying interactions has given us a new systematic strategy. We expect now to connect molecular architecture with molecular recognition and assembly. For example, the very strength, flexibility and adaptability of the collagen "rope" can come to be seen as a consequence of the way fibers are held together by newly identified forces.

Future Plans/Trends

Implementation of this strategy requires further force measurements in protein systems such as fibrous proteins, model peptides and oligopeptides. Some of this work has already begun. Currently we are preparing samples from model triple helical peptides for future force measurements. Another part of this strategy is measurement of the effect of chemical or genetic modification of protein surfaces on intermolecular forces. Chemical modification of the suspected recognition sites on collagen is the next planned step in this direction.

Formation of highly ordered collagen films with controlled interactions between protein filaments is a promising way to develop new implants for reconstructive surgery. We are planning to start testing some of these new materials together with Dr. K. Salyer from Humana Advanced Surgical Institutes, Medical City, Dallas.

Theoretical work will concentrate on understanding the structural changes induced on interacting surfaces. One of our strategic goals is the development of realistic potentials from directly observed intermolecular forces. These can be later

combined with the exactly solved theoretical models to form a new basis for computer prediction of molecular structure and interactions.

Forces Between Collagen Molecules; Collagen Fiber Assembly

Forces between collagen triple helices have been measured in various sodium salt solutions. Both the equilibrium separation at zero applied stress and measured magnitude of the repulsive force at low pressures decrease with increasing salt concentration. Salt does not appear to act on the intermolecular forces through screening of the electrostatic double-layer repulsion. Rather, salt is preferentially excluded from the space between collagen helices so as to apply an extra osmotic pressure. This finding has been confirmed by comparison of intermolecular spacings in NaCl and in solutions of large polymers of known osmolality. It appears that only 33% of NaCl penetrates between the helices. The osmotic stress applied by physiological salt solutions might play an important role in preventing the collagen fibers from overswelling and losing their integrity. The forces measured between collagen helices are a combination of an exponential short-range repulsion and a longer ranged attraction responsible for spontaneous assembly. As we have previously shown, from 5° C to 35° C the relative contribution of the attraction to the net force increases with temperature.

We have now demonstrated that as pH is reduced from 7.5 to 6 or lower, both the attraction and the temperature sensitivity are completely removed. The same effect has been observed upon addition of glycerol into the bathing solution of pH 7.5. However, at pH 6, when the attraction is already removed, the addition of glycerol has no effect on intermolecular forces.

These results practically rule out the hydrophobic effect, usually invoked to explain temperature-favored assembly of proteins. They argue against the electrostatic nature of the attraction as well. The spontaneous assembly of collagen fibers at pH 7.5 appears to involve formation of water-mediated hydrogen bonds between the apposing polar residues. This exponential attractive hydration force, obtained by subtraction of forces measured at pH 7.5 from the purely repulsive interaction at pH 6, is in good agreement with theoretical predictions.

Impact/Value

Building on these observations, we have developed new theoretical models of molecular recognition and assembly. This is intended to be the beginning of a practical vocabulary of forces to be incorporated into computer algorithms for protein folding, contact, and ligand or drug binding.

Modification of Forces Between DNA Molecules by Small Solutes

Forces between DNA molecules have been measured in solutions of methanol, glycerol, ethylene glycol, glucose, sucrose, and sorbitol. These electrically neutral solute molecules are small enough to penetrate into the space between DNA helices. Methanol condenses DNA, the effect seen as significant reduction in the measured repulsive force. Glycerol and ethylene glycol have almost no influence on the DNA spacing. Sugars induce extra swelling of DNA as if they produce an extra repulsion. A particularly strong effect is seen in the presence of sorbitol.

These results can be interpreted as an extra osmotic pressure produced by small solutes. These act either from within or from outside of DNA fibers, depending on whether the solute is preferentially included or excluded from the space between DNA molecules. The increased or reduced concentration of the solute inside the fiber is apparently due to the interaction with DNA surfaces. The inferred force between the solute molecules and DNA is exponential and appears to be proportional to the solute size.

Impact/Value

The adsorption vs repulsion of small molecules from a macromolecular surface is a critical factor in

stabilization or destabilization by small-molecular-weight species. Questions of denaturation, of polymerization and of switching between molecular forms require the kind of information revealed by the action of small solutes on intermolecular forces.

Chain-Melting Re-entrant Transition in Diacylphosphate Bilayers

In the course of measuring forces between di-dodecyl-phosphate bilayers, which are being used to study the fusion of such membranes, unusual re-entrant chain order transitions have been detected by our collaborator, Dr. R. P. Rand and his coworkers at Brock University, Ontario, Canada.

The lipid chains freeze and then almost immediately melt again when the bilayers are pushed together. These transitions occur when the bilayers are still separated by as much as 30 to 140 angstroms. We have suggested a theoretical model of this phenomenon based on the balance among (a) electrostatic energy of bilayer interactions, (b) the elastic compressibility energy within bilayers, and (c) the work of the osmotic stress. This theory explains the separations, unusually large for osmotic-stress-induced chain-order transitions, and the observed salt dependence of these transitions.

Impact/Value

Rearrangements of membrane lipids through membrane-membrane interaction are central to the process of membrane fusion. The strains within lipid bilayers, seen through phase transitions, might modify the behavior of membrane proteins (such as was noted in our lab last year with the influence of lipids on channels formed by alamethicin peptides).

Osmotic Pressure of Ordered Colloidal Suspensions

The osmotic stress technique for intermolecular force measurement has been extended to observe ordered suspensions of charged phospholipid and microsomal vesicles. The control measurements have been done for suspensions of charged latex spheres of known diameter.

Impact/Value

Such suspensions are an excellent model for organization under crowded conditions such as occur inside a cell. They provide an excellent vehicle for systematic physical analysis of molecular ordering and for design of practical assembly systems.

Noninert Glue in the Surface Force Apparatus

During the past several years, the surface force apparatus has been widely regarded as a means to measure forces between macroscopic surfaces. We have found that the glue commonly used in the surface force apparatus is not inert, but rather creates a measurable amount of water-soluble material that is also surface active. Should there be significant contamination, many of the conclusions drawn using this technique will have to be re-evaluated.

Physics of Ionic Channels and Other Proteins with Aqueous Cavities

The purposes of this initiative are:

- To delineate structural features of ionic channels by their reaction to polymers of varied size
- To observe channel kinetics through physical "noise" and to determine rapid events such as binding and unbinding of protons from ionizable sites
- To test for the channel-forming capabilities of antibiotics known to perturb cell membrane transport
- To relate forces measured between macromolecules to the energies that drive channels or proteins in solution between functioning states of different structures.

Ionic channels are reconstituted into bilayer membranes and electric current kinetics are studied as functions of applied voltage and osmotic stress.

Direct Observation of Proton Binding and Unbinding at Ionizable Groups on a Protein Surface

With "patch-clamp" and "blm" membrane channel reconstitution techniques, one becomes spoiled, taking for granted the ability to observe one protein molecule or one channel. But it is possible to go one step further, to watch one ionizable group bind and unbind a proton, and to measure the on- and off-times of molecular association.

It was possible to measure the binding and unbinding rates of a proton inside an ionic channel. This was done by frequency analysis of the electrical "noise" created by proton binding. One sees that the responsible ionizable sites have a pK of 5.8 and that the association and dissociation rate constants are 8 x 10^9 and 10^5 sec^{-1}, respectively. Our experiments demonstrated the possibility of studying chemical reactions in a single microscopic (in fact, nanoscopic) "cuvette" in which only several molecules participate. We suggest that this approach will prove useful as a new powerful tool for determining functional structure of channels. It might also provide ways to look at proton fluctuations on proteins in solution with immediate consequences for the way we think about protein fluctuations.

Probing the Dimensions of Channel Proteins

Contrary to expectations based on consideration of increased viscosity, alamethicin channels current "bursts" speed up in the presence of water-soluble polyethylene glycols (PEGs) and dextrans. Added polymers reduce the probabilities of transition to higher conductance states, but do not change channel lifetimes. They thereby shorten the duration of current "bursts."

These modified probabilities and kinetics reveal the action of polymer osmotic stress to suppress channel formation. The osmotic action of large, fully excluded polymers shows that some 100 water molecules are taken up by the channel from the solution upon each transition to an adjacent higher conductance state.

Small polymers are seen to enter ionic channels. The partial osmotic action of different-size polymers reveals the extent of their exclusion. One can relate the degree of each polymer's exclusion to its known size and consequently to the radius of the channel pore.

This strategy introduces a new method for interrogation of ionic channel structure using water-soluble polymers. It also opens up a new way to study the statistics and energetics of soluble polymers entering cavities of well-defined size.

Channel Formation by Soluble Antibiotics

Novobiocin has been found to form ionic channels in lipid bilayers. This is an entirely unexpected property of this aromatic nitrogen-containing antibiotic often used as a pharmacological agent to enhance the responses of sodium-specific, amiloride-sensitive nerve fibers to sodium chloride. We have found that it also forms ion channels in lipid membranes, suggesting that its ability to act as a salt enhancer may be due to cation-selective channel formation in cell membranes. The type of fatty acids composing the phospholipids used to make the host membrane does not matter, but phospholipid charge is important. Negatively charged lipids allow formation of higher conductance states not found in neutral lipids. Recognition of channel-forming capabilities, at least in this one case, reveal new possibilities for visualizing modes of antibiotic action.

Significance

Channels can be used as rapid detectors of individual molecular events, in particular of proton binding to ionizable sites. Polymers of different sizes can be used to gauge channel dimensions and elucidate changes in structure that accompany changes in function. Diseases attributable to defective ionic channels may be approached through these new techniques of probing channel mechanics.

Work Plan

We will apply the osmotic stress of excluded and partly excluded polymers to several different ionic channels, probably including ion-specific channels from nerve membranes. We hope to use channels of known dimensions to study polymer-cavity interactions to see how polymer conformation, such as is seen in radius of gyration, will determine partition into small spaces. We would like also to continue work on proteins in solution, particularly the allosteric transition of hemoglobin and the helix-coil transitions of oligopeptides.

Richard J. Feldmann

Modeling the Mechanism of Protein Folding

The protein-folding problem remains one of the central unsolved problems in molecular biology. Approaches to this problem generally fall into one of three categories: 1) molecular dynamics which simulates every atom motion, 2) abstract models which use an abridged representation of each amino acid, and 3) secondary structure prediction methods which use only the name and sequence position of each amino acid. The computational requirements for the first approach, molecular dynamics, make it useful only for studying the folded state. Secondary structure prediction methods, the third approach, all hit a barrier which limits them to approximately 62% prediction accuracy. In choosing the middle ground of the abstract model, we have added a topographic component which represents the solvent and counter ion environment of the protein. The goal of this work is to produce a Fortran program which will calculate the three-dimensional structure of any protein to within 2 angstroms in a day of computing on any modern workstation.

In the preceding years, a collaboration with Drs. J. David Rawn and George S. Michaels developed a topological model of a protein and its environment. This year was spent implementing, testing and modifying this model. At first the results were very encouraging. We used Triose Phosphate Isomerase (TIM) as the test protein because it has a strong stable folding pattern. Crystallographers have already solved the structures of more than a dozen independently evolved TIM-like proteins. The simulation program formed helices where they occur in TIM and nonhelical regions where beta strands occur. The helices emerge from the hydrophilic component of our model but the beta barrel, which in TIM is formed from eight beta strands, could not be formed. We tried adding many different features to our topological model as well as making many computational experiments in which we varied the parameters of the model. We added features to the program which permitted comparison with the crystal structure.

As the year proceeded, the topological folding simulation program was moved from the Apple® Macintosh®, on which the program development is regularly done, to a number of different brands of computers (IBM® RS-6000, HP-730, DEC® Alpha, Convex, Cray, Intel® IPSC, Fujitsu) in an attempt to obtain more computing time. We found that collections of workstations, or "farms," were the most cost-effective source of bulk computational power. The "farms" of IBM® RS-6000 workstations in the division (6 machines) as well as at NASA Lewis (32 machines) and Argonne National Laboratory (128 machines) proved to be the most useful to our work.

At first, bulk computing seemed to be the cure for our problems, but in the end, it became clear that our model had to be changed in a more drastic manner. For several decades scientists have believed that both hydrophobicity and hydrophilicity of an amino acid sequence are responsible for the specificity of the protein folding. In our topological model of a protein we sought a balance between the roles of the hydrophilic and hydrophobic aspects. We developed a new computer graphic representation of the structure of a protein. The protein is represented as a circle with arcs representing the hydrogen bonds and hydrophilic loops in one diagram and hydrophobic loops and bonds in another diagram (Figure 8). We developed a novel approach in which PostScript® files can be generated directly from any

Fortran program. Once the graphic was developed, we applied it to many different proteins and to all of the folding states of TIM. These graphics showed that the hydrophobic bonds dominate the progress and direction of protein folding. Rather than a balance between only two forces, we now see protein folding as a hierarchy of bond positions and life-times. Hydrophobic bonds have an indefinite life-time but a

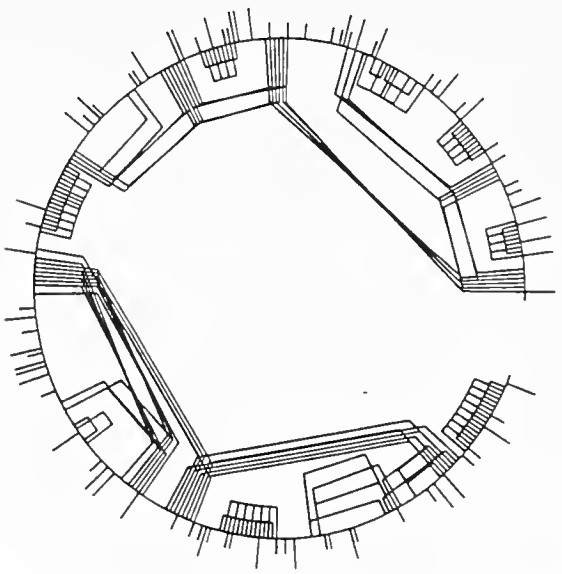


Figure 8. Diagram of the hydrogen bonding in glyceraldehyde-3-phosphate dehydrogenase showing the alpha helices and the parallel and anti-parallel beta strands.

very high positional mobility. Hydrogen bonds, being six times less prevalent in TIM than the hydrophobic bonds, have a finite life-time but a fixed position. Salt bridges which are sequence specific are relatively few in number, have a rather short life-time, but act at critical junctures in the folding of the protein. Disulfide bridges are very few in number, are sequence specific, and have very long life-times. The different bond types serve to pin the protein together at different times and places. Our graphic representation development has helped us to see the importance of this hierarchy of bond types.

The crystal structure of TIM for every length was analyzed using the circle graphic. This is equivalent to looking at the putatively best fold for every step of ribosomal synthesis of the protein. For any length of protein there will be one distribution of hydrophobic bonds. When the next amino acid is added to the protein, the hydrophobic bonds shift slightly. This shifting of the hydrophobic bonds is based on an algorithm which considers the number of hydrophobic bonds that each amino acid type can sustain and the distance between a pair of amino acids. On average, two hydrophobic bonds are added per amino acid, but an additional three change their positions. In our simulation, as the protein is synthesized, the distribution of hydrophobic bonds is modulated. Pairs of large hydrophobic amino acids (like phenylalanine and tyrosine) act to "store" hydrophobicity. Later in protein synthesis their hydrophobicity may be redirected to amino acids quite distant in the sequence. This can produce the tertiary structuring which the secondary structure prediction methods miss. The small hydrophobic amino acids (like glycine and alanine) serve to regulate the flow of hydrophobicity within the folding protein in ways that we still do not fully understand. Amino acids with a medium amount of hydrophobic capacity (like valine and leucine) and in some cases hydrophilic residues (like serine, lysine and glutamate) are pivotal to the interplay between the hydrophobic and hydrophilic aspects. Analysis of the circle graphics for several different classes of protein architectures has shown that, in all cases, similar coherent modulation of hydrophobic bonding exists as a function of protein length. Despite the apparent dominance, in our model, of the hydrophobic aspects of a protein, the specificity of the hydrophilic aspects of many amino acids must still be incorporated into our concepts of protein folding.

Future Plans/Trends

We believe that the prognosis for solving the protein-folding problem is quite good at the moment. Even though 40% of the Fortran code so laboriously developed over the past year had to be discarded, it

is only the central core of our model which has to be rewritten. The pattern of insight, formulation, implementation, testing, analysis, despair and then insight will be carried forward until this problem is solved. The problem has proven to be more difficult than we had estimated a year ago. The key insight of this year is that the modulation of the position of the hydrophobic bonds can contribute to secondary and tertiary folding specificity. What remains is to make the simulation program do this in a reliable manner for any protein sequence.

Impact/Value

We hope and expect that the subtle interplay between thinking, code writing and computing will eventually produce a set of concepts and rules which will provide a general, reliable and rapid protein-folding algorithm. Since the mechanics of data files and bulk computation are all in place, we shall concentrate on the development of our algorithms.

Molecular Graphics and Simulation Section

Bernard Brooks, Ph.D., Head

Molecular Dynamics Simulations of Biological Macromolecules

The Molecular Graphics and Simulation (MGS) section studies problems of biological significance using the theoretical techniques of molecular dynamics, molecular mechanics, modeling, *ab initio* analysis of small molecule structure, and molecular graphics.

Research Involving HIV Proteins

The MGS, in collaboration with the Biophysics Laboratory of the FDA Center for Biologics Evaluation and Research has been part of the NIH Intramural HIV Targeted Antiretroviral Program since 1987. Current studies include:

- theoretical analysis of inhibitor binding to the active site of HIV-1 protease using molecular dynamics and free energy perturbation approaches
- HIV-1 protease cleavage of viral polyproteins: molecular dynamics investigation of the chemical mechanism.

The primary goal of these studies is to elucidate the mechanism by which HIV-1 protease binds and cleaves viral polyproteins. The cleavage reaction is a necessary step in the maturation of the HIV-1 virus. Thus HIV-1 protease is a possible target for AIDS therapies, and it is the object of intense theoretical and experimental study. An understanding of the mechanism of reaction would be of great value in the search for effective inhibitors for the protease. Secondary goals of this study include the development of new algorithms for the investigation of complex reaction processes and insight into the design of inhibitors.

This study uses a combined quantum mechanical and molecular mechanical (QM/MM) potential interfaced with the molecular modeling program CHARMM in order to study the protease/substrate system. The QM/MM potential allows the simulation of bond breakage and formation. The study comprises several steps: (1) the selection of proposed reaction mechanisms, (2) the determination of reaction paths and their energetics for each mechanism by appropriately searching the potential energy surface, and (3) free-energy perturbation simulations along the reaction paths to provide entropic information. The profile of free energy change along the reaction paths will determine the most likely reaction mechanism. The reaction mechanisms selected for study are a general acid/general base mechanism, a nucleophilic attack, and a mechanism involving a zwitterion. A suitable x-ray crystallographic structure of an inhibited complex of HIV-1 protease has been chosen, and a low-energy structure with which to begin reaction-path determination has been generated by simulated annealing of the x-ray structure. A method for determining reaction paths for processes involving large molecules, the conjugate peak refinement method, has been tested. Several of the reactants, products, and intermediates for the proposed mechanisms have been minimized with the QM/MM

method to define endpoints for the reaction paths. Methods of visualization, which is critical to this project since the choice of reaction coordinate requires considerable physical insight, have been developed. Preliminary simulated annealing studies suggest that the oxygen on the scissile carbon may be favored. It is our goal to provide theoretical evidence favoring one of the proposed mechanisms of the HIV-1 protease cleavage reaction.

Other Applied Research on Molecules of Biomedical Interest

Applied simulation research uses molecular dynamics simulations to predict function or structures of peptides and proteins, often with application to specific biomedical goals, such as vaccine or therapy development. Specific studies include:

- structural characterization of a heme-myoglobin adduct using molecular mechanics
- simulation and modeling of intermediate filament (IF) proteins
- identification of peptides which bind to human major histocompatibility complex (MHC) DR1.

In collaboration with the NHLBI, we have used molecular modeling techniques to study a heme-myoglobin adduct in which the covalent structure of myoglobin has been altered, functionally changing myoglobin from an oxygen storage protein to an oxidase. The role of solvent in the relationship between protein structure and function was evident in that the active site of the modified structures was much more accessible to water molecules than that of native myoglobin.

Mutations in intermediate filament (IF) proteins are implicated in keratinizing disorders of the skin. The goal of the simulation and modeling of IF proteins is to elucidate the molecular-level effects of these mutations.

Basic Research

Basic research provides a better understanding of biochemical systems. Emphasis has been on simulations to analyze structure – function relationships and other properties of macromolecules. Simulation results are compared with experimental data whenever possible. This type of research is also needed for the testing and evaluation of new methods and models. Specific projects include:

- the effects of temperature on protein dynamics
- the effects of hydration on protein dynamics
- molecular dynamics simulations on staphylococcal nuclease: comparison with NMR data
- harmonic analysis of large systems
- modeling and simulation of the lipid bilayers in crystal and gel phases
- molecular dynamics simulation studies of DNA: the B-Z junction
- examining long-range deuterium isotope effects in C-13 NMR spectra
- solvent induced forces between two hydrophobic groups
- an examination of internal friction in proteins.

By incrementally increasing the number of explicitly included water molecules, we have studied the effects of hydration on carboxy-myoglobin using molecular dynamics simulation (see figures on front and back covers of this report). In agreement with experiment, our simulations indicate that myoglobin is effectively fully hydrated by 350 water molecules. The large body of simulations provides an atomic-level description of the hydration shell and gives insight into hydration's influence on protein structure and function.

Long-range deuterium isotope effects in C-13 NMR spectra has been investigated with *ab initio* calculations for deuterated and undeuterated binuclear aromatic compounds. These very small effects (in ppb) were measured by collaborators Dr. Edwin Becker and Dr. Vikik-Topic from the Laboratory of Chemical Physics, NIDDK, and the resulting theoretical predictions correlate very well with experimental data.

Impact

These projects provide two important benefits. First, they provide insight at the molecular level to complex processes and phenomena, and have the

potential to impact the design of effective therapies. Second, the difficulties and deficiencies encountered in pursuing these projects drive the development of new methods, so as to facilitate future molecular simulations.

Future Plans

In FY94, the MGS will continue to study relationships between structure and function and develop the theoretical analysis of inhibitor binding, specifically studying the mechanism of the HIV-1 protease. Several new projects will be initiated, including:

- simulation and modeling of HIV-1 reverse transcriptase (RT)
- the temperature dependence of the behavior of extreme thermophile proteins
- structural analysis of T4 lysozyme mutants
- reaction path modeling of DNA photolyase using quantum mechanics/molecular mechanics (QM/MM) methods.

In the immediate future, we propose to look at the RT system, examining overall protein behavior and structure. This work will be based on the recent low resolution structure of reverse transcriptase from the Steitz laboratory at Yale. The initial study will focus on structural issues, such as energetics, stability and strain. Longer term goals involve examining the interaction of RT with the membrane and its interactions with RNA and DNA. Realistic simulations of the membrane environment make it now feasible to examine membrane-bound proteins at the atomic level without introducing effective potentials for the lipid interface region. Another longer term goal will be to examine the mechanism of RT using a combined QM/MM investigation.

Development of Theoretical Methods for Studying Biological Macromolecules

Algorithms and Software

New theoretical techniques under development are often coupled with software and hardware development. These involve the generation of new simulation techniques and the systematic testing and evaluation of methods. Projects include:

- development of quantum mechanical potentials and appropriate algorithms for use in molecular dynamics simulations
- determination of protein structure by NMR and molecular modeling
- slow growth homology modeling (SGHM) for the determination of protein structures
- development of an optimized protocol for the preparation of low-temperature states
- development of flexible molecular dynamics techniques that remove high frequency degrees of freedom
- free energy perturbation simulations in solution, examining the effect of restraints
- conversion of physical models into three-dimensional coordinates for computer analysis and simulation
- development of ray-traced molecular graphics software for Hewlett-Packard™ workstations, high-resolution color printers, and for movies using National Television System Committee (NTSC) video equipment
- adaption of an efficient Newton minimization procedure for CHARMM and biomolecular applications.

One major advance is the combination of the large *ab initio* software package GAMESS (General Atomic and Molecular Electronic Structure System) and molecular mechanics program CHARMM. This allows the study of critical portions of a macromolecular system at a high level of accuracy. We are developing software so that molecular mechanics calculations using CHARMM can be combined with QM calculations at several levels of exactness of theory.

Our research continues to focus, in part, on optimizing protocols for the simulation of biomolecules. Among the important issues being addressed are the accurate treatment of solvent effects, the efficient approximation of long-range forces, and the appropriate preparation of low-temperature states.

Analysis of simulation of protein dynamics often demands visual representation. For a movie of only few minutes in duration, several thousand picture frames need to be generated and stored on video equipment. High-quality rendering requires that ray-traced graphic images be generated and stored. Software has been developed which enables this procedure to be performed without human intervention, so that movies can be made overnight and on weekends and stored on high-quality optical disks.

A software package has been developed which allows three-dimensional coordinates to be extracted as pictures. An interface with a camera connected over a SCSI interface to a workstation allows the user to manipulate images so that three-dimensional coordinates for complex models can be obtained from stereo images of plastic physical models.

Parameters and Models

Parameter sets and models are generally available for most macromolecular systems, but there is considerable room for improvement, and alternate models that improve realism or reduce computational costs need to be examined. This effort involves the refinement of parameters and the exploration of alternate energetic models for molecules and environmental conditions. Ongoing projects include:

- development of parameters for alkane systems
- approximation of long-range interactions in macromolecular simulation variants of the cell multipole method
- new methods for long-range truncation of the energy potential
- evaluation and comparison of implicit and explicit water models for simulations examining the hydration of proteins
- molecular dynamics simulation studies of DNA: analysis of the parameter sets using an infinite DNA helix
- analysis of conformation in response to changes in solvent, *ab initio* studies.

We have undertaken a comprehensive evaluation of spherical cutoff methods for truncating long-range electrostatic interactions. Both traditional approaches and new methods developed in our lab have been surveyed and a checklist of desirable features has been proposed. A detailed comparison of the many cutoff schemes based on simple test cases and on simulations of hydrated myoglobin has been generated.

Development of good methods to simulate macromolecular behavior in a solution is still a problem. Recently, implicit methods using atomic solvation parameters became popular. We have developed software that runs on parallel computers to study the effects of these implicit solvent models. It has been shown that current implicit methods and parameters are inferior to available models involving explicit treatment of water.

Impact

Development of new methods, models and parameters is essential for the future of macromolecular simulation and modeling.

Future Plans

In FY94, the MGS will continue a broad effort to develop new methods. Methods to improve the accuracy from free energy perturbation simulations by the development of a new integration procedure for molecular dynamics will be explored. Also, methods for treating solvent implicitly to provide for hydrophobic effects without the explicit inclusion of many water molecules, and methods to properly treat electronic polarization in molecular dynamics simulations will be explored. These methods will be applied to a variety of macromolecular systems. These will include proteins and substrates from the HIV-1 virus, heme proteins such as myoglobin, interleukins as well as small peptides. New projects include:

- further refinement and examination of free energy techniques

- development and use of a polarizable and flexible water model
- three-dimensional structure determination of proteins from a simplified topological description (with Richard Feldmann).

Development of Advanced Computer Hardware and Software

With the advent of new computer technology amenable to large-scale scientific computing, software and hardware development efforts are essential for optimal use of these resources. The efforts include developing techniques to exploit parallel multimachines, writing assembler code for optimal performance on commercial processors, and establishing parallel workstation clusters for high-efficiency simulations at low cost.

Massively Parallel Computers

Development of methods and software to make productive use of parallel MIMD machines for use in macromolecular simulations is under way. The initial global communication approach has been successful in providing an efficient full-feature version of CHARMM. This parallel version of CHARMM has been extended to run on almost any MIMD parallel computer platform: Intel® iPSC/860, Intel® delta, Thinking Machines CM-5, IBM®/SP1, and on clusters of workstations. Current projects include:

- a scalable molecular dynamics algorithm for massively parallel machines and large workstation clusters
- development of parallel QM/MM methods
- development and efficient use of a high-speed cluster of HP735 workstations
- further development and support of CHARMM.

Our current development effort involves a scalable algorithm that promises to greatly reduce the communication cost for very large MIMD machines or for large workstation clusters. The nature of an ideal scalable algorithm is that the time spent for communication is actually reduced as the number of nodes increases. In the algorithm that we are adapting, the communication costs scale as the reciprocal square root of the number of processors.

Workstation clusters provide a highly competitive environment in terms of cost performance for macromolecular simulations. A workstation cluster based on the HP730s has been assembled. Parallel software has been developed and evaluated as a function of network connectivity (Ethernet, Token ring, or FDDI). The initial phase of this work was conducted in collaboration with Dr. Robert Martino and Stan Erwin of CBEL, using the DCRT Intel® 128-node processor. The communications routines were based on the work of Dr. Robert van der Geijn.

Impact

The parallel version of CHARMM developed at NIH is being used on many MIMD machines and it has gained widespread acceptance among CHARMM users. This full-feature version of CHARMM enables MIMD technology to be put to practical use for molecular dynamics. The parallel version of CHARMM is now being used for most of the research projects in the MGS, and it is proving to be reliable.

Future Plans

In FY94, we hope to complete the workstation cluster by upgrading all of the nodes to HP735/755s and to enhance the communication speed by the acquisition of an asynchronous transfer mode (ATM) switch. To enhance I/O capabilities and availability of virtual memory, a large disk will be added to each node. This cost-effective cluster of 16 workstations should, in theory, perform at the level of a four-processor Cray Y/MP for macromolecular simulations with an efficiency of roughly 85%. The new scalable parallel algorithm will be evaluated and put to use on this cluster and on other highly parallel systems.

LSB Support Activities

The MGS is actively supporting molecular modeling and simulation needs at the NIH, both through consulting and formal training. Direct services provided by the MGS unit include:

- research support and guidance for NIH scientists
- provision of an NIH resource for short-term graphics and modeling needs
- support for software packages on a variety of hardware platforms
- examination and evaluation of new hardware
- assessment of needs at NIH and provision of policy recommendations to DCRT management and other NIH organizations
- assistance to other DCRT sections in making their computational resources useful for the research needs of NIH.

Courses and Seminar Series

The MGS supports four courses which are given periodically:

- CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics
- Usage and Applications of Molecular Quantum Mechanical (QM) Programs
- Molecular Dynamics for Problems in Structural Biology
- Molecular Graphics: Creating Pictures and Videos.

The MGS is also conducting an active seminar series for computational chemistry, and it conducts a book review series, both of which are open to interested scientists.

Future plans

MGS will continue to be a resource for NIH, provide direct collaborative assistance, and give courses and organize seminar series and book review series.

Analytical Biostatistics Section

Peter Munson, Ph.D., Head

Statistical Methods for Molecular Biology, DNA and Protein Structure/Function

Purposes and Goals

New opportunities for large-scale computations have stimulated the development of many new statistical techniques (e.g. Bootstrap, cross-validation, Expectation-Maximization algorithm, Monte-Carlo, projection-pursuit regression, neural networks). The purpose of this project is to investigate the applicability of modern statistical methods to problems in molecular biology and in particular, structure/function prediction from linear DNA or protein sequence data. Further, we seek to modify and extend classical statistical procedures, such as maximum-likelihood, where appropriate in this context. The goal of this approach is to provide, adapt and apply methods optimally suited to this family of problems.

Methods

Primarily, the methods used are those of mathematical and applied statistics. From the characteristics of existing and anticipated data sets and the nature of the research questions, appropriate statistical methods are explored, both from a theoretical and computational viewpoint. Optimization algorithms such as maximum-likelihood or simulated annealing are used to find best parameter sets. Simulation may be used to characterize the properties of a method, as distinguished from its performance on a single data set. We used cross-validation and calculation of the effective degrees of freedom to control for the dimensionality of the model. Penalized log-likelihood methods were developed to reduce this dimensionality effectively. Kernel density estimation techniques were used to predict secondary structure nonparametrically.

Major Findings

In the context of protein secondary structure prediction, we have completed the development of a quadratic-logistic prediction model. This model considers that small sequence fragments within the protein chain should determine the secondary structure (alpha helix, beta strand or random coil) of the residue in the center of that fragment. It would be impossible to estimate the parameters of arbitrarily complex models relating structure to sequence. Therefore, we begin with models using only the first-order (linear-logistic) and second-order "quadratic-logistic" terms. We anticipate that such models would capture much of the important biophysics of the problem, especially the "potentials" relating preference of residue pairs to be situated next to each other in the folded protein. While other investigators have previously considered the effect of pairwise interactions, they did not use optimally efficient statistical methods, nor did they adequately control for the effects of overparameterization. In the full pairwise model, there are over 100,000 parameters to be estimated, which far exceeds the size of the available dataset. Our approach exploited the periodicity of both alpha helix and beta strand, to reduce the number of effective parameters. Further reduction was obtained via a penalty term in the log-likelihood optimization. With this method, we slightly improved the prediction accuracy of earlier methods. We then improved the prediction accuracy further with use of the quadratic terms (up to 65.9% correct). In this, we estimated all 400 pairwise residue preferences for alpha helix or beta strand. Finally, we showed that the database effectively limits models of this form to about 800 effective parameters. A consequence of this idea is that further improvement in secondary structure prediction is unlikely until the database grows substantially, and that with such growth, the prediction accuracy could rise to as much as 74%, the prediction score without crossvalidation.

We have also built a nonparametric discriminant function model using kernel density approaches. The sequence data were mapped into a continuous metric space, then the space was automatically constructed using multidimensional scaling. This approach also attained a similar prediction rate (about 64%) for a model with optimal bandwidth. Since this nonparametric approach can theoretically fit models with arbitrarily high complexity, our result suggests that database size (and also tertiary interactions within the folded protein) limit prediction accuracy. Thus, it is not the complexity of the model, *per se*, which limits prediction accuracy. In this study, we have also shown that protein class estimation from sequence can possibly add 3-4% to the overall prediction accuracy.

We are also investigating the potential of alternative graphical representations of proteins for visual classification and understanding of the "space" of protein structures. We have used 15-angstrom "contact" maps to represent the gross structure of the main chain of the proteins in the Protein Data Bank. These two-dimensional maps of the three-dimensional structures fall naturally into classes, based on their size and visual texture. Structural motifs (alpha-alpha interactions, beta-beta interactions, alpha-beta interactions, turns) are easily identified in these maps. An organized "atlas" of protein structures in this representation has been prepared.

In a separate project, we have identified the longest known DNA sequence over which long-range correlations of base usage are apparent. The recently sequenced yeast chromosome III displays these correlations up to the 64-Kbase range. We also developed a statistical test to determine if these apparent correlations could be due to random fluctuations or even to fluctuations implied by correlations at shorter ranges, as might arise within a single gene or group of genes. For chromosome III, the correlations are statistically significant out to about 8 Kbases. It is still unclear what mechanism produces these correlations, as it would be surprising to find any DNA transcription or translation process which might effectively extend over such a large range. Current suggestions include the restrictions imposed by DNA packing in the nucleus, DNA attachment of membranes, and an evolutionary

process involving gene duplications followed by mutation. The pervasiveness of the long-range correlations suggests that a very general biological process may be involved.

Impact/Value

Molecular biology, and DNA sequence data in particular, is one of the fastest growing data sources in biology today. Improved statistical methodology is required to deal with research questions which arise in this context. Currently, protein structure and function prediction from sequence is a major problem whose solution would have dramatic impact on the utility of sequences generated in the Human Genome Initiative. Given the long-standing nature of this problem, it is clear that even incremental improvements on current methodology will ultimately prove valuable to the biological and medical communities. Optimal statistical methodology should be able to provide those improvements, and should provide better answers to what can and cannot be said about new sequence data as they are generated. Ultimately, new statistical methodologies are needed both as data sources increase and as biophysical principles governing protein structure and function are established.

Proposed Course

We will continue to enhance models of secondary structure prediction using additional data sources, and incorporating the dependent nature of structural state formation. We will investigate Markov and related state-space models for this purpose. Also, we will seek better automated structure classification methods, and investigate their role in structure prediction from sequence. We will also consider empirical potential models in view of secondary and tertiary prediction.

Statistical and Computational Methods for Physiology, Pharmacology and Endocrinology

Purposes and Goals

We develop, test, apply, and disseminate statistical and computational methods for studies in physiology, pharmacology, endocrinology and related areas. Our aim is to advance these studies by exploiting optimal statistical methodology, and make these technologies widely available through distribution of program packages, teaching courses, and developing program manuals, as well as individual consulting. Where necessary, we seek to develop new statistical methodology which exploits the computational facilities now generally available to many laboratories.

Methods

We use mathematical statistical theory and methods to refine classical procedures, develop novel approaches, and characterize their statistical behavior. Modern computationally intensive statistical approaches often require the unique resources available at DCRT. Where appropriate, we package such procedures for distribution and use by investigators in the laboratory, and develop, maintain, document and distribute software code.

Major Findings

A prominent study (Science 1992; 259:801) recently claimed that human growth in children is "saltatory," occurring in spurts separated by quiescent periods lasting up to weeks. While this hypothesis fits many preconceived notions of parental observers, the published data do not really support the claim when closely inspected. The statistical problem arises because the "saltatory" hypothesis is really a growth model with many parameters (the length of each quiescent period) while the continuous growth model really has only one (the constant growth rate). Therefore, it should be no surprise that a data set can be produced which

appears to better fit this more complex model. The relevant question should be: given the additional flexibility of the saltatory model, does it fit the available data significantly better? In collaboration with NICHD investigators, we analyzed two sets of data, one in animals, where daily growth can be precisely measured, and another in human infants, to determine whether the continuous growth model was satisfactory, and if the saltatory growth model could be rejected. Analyzing the daily growth velocities, neither data set showed any suggestion of a bimodal or composite distribution which would be characteristic if growth occurred on only a small fraction of the days. Moreover, there was no correlation between growth velocities measured by separate indicators (weight, leg length, head circumference, body length) as would be expected if saltatory growth were present. With most measures, it was possible to reject the saltatory model completely, if a fairly narrow definition of saltatory growth was adopted (growth on less than 8% of days, zero growth otherwise).

A computer program (LIGAND), developed within this section and widely distributed, encompasses a nonlinear least-squares analysis of ligand binding studies. This program was significantly enhanced for use on the Macintosh® computer, and now allows for flexible specification of model constraints, and fully general multiple-ligand study designs. More than 150 requests for copies of this program were satisfied. An enthusiastically received half-day short course in the use of this program was given. A second program (ALLFIT) continues to be popular. This program has also been adapted to the Macintosh®, and is now capable of convenient dose-interpolation for assay applications. Several consultations with NIH and outside investigators were performed on problems relating to these programs.

Neural networks represent a useful new tool in computer science, whose importance in the solution of practical problems is only now being widely appreciated. In addition to sponsoring an NIH-wide journal club on this topic, the section has begun several investigations into the statistical properties of artificial neural networks. With an NHLBI investigator, we are looking at the potential for networks to recognize the time-varying spectral signature of several compounds. In another study, we are seeking to understand the source of the improved prediction of mechanism of action of several drug compounds, using neural networks with hidden layers. The improvement in prediction may be due to coding considerations within the network or to regularities within the dataset "discovered" by the network.

Impact/Value

Statistical method development will continue to have practical value to both experimentalists and theoreticians. As a practical matter, it is extremely valuable to "package" new methodologies in a form accessible to investigators in the target discipline. Emphasis on the development and distribution of computer programs facilitates this technology transfer, and can ultimately improve the quality of investigations done in an entire discipline. The theoretical value of "methods research" is appreciated within a narrower segment of the scientific community. DCRT's commitment to provide excellence in scientific computation support to the NIH Intramural Program is well served by this and other methods-development efforts.

Proposed Course

We will continue to expand the research effort into the applicability of new statistical methods such as artificial neural networks and nonparametric density estimation methods. Both of these approaches promise wide applicability in divergent areas of scientific study. Support, development and distribution of several computer programs will continue.

George Hutchinson, Ph.D.

Adaptive Computing and Biomedical Application

Objective and Goals

The project objective is to develop adaptive methods for computing and for developing computer software, and to apply them to selected biomedical research problems. Methods include artificial neural networks and related approaches and the use of advanced computer languages to permit rapid implementation of mathematical models. Goals include improved understanding of adaptive methods and development of applications which are useful for problems of biomedical research.

Methods

Research continued on the formulation of algorithms for training neural networks with unequal error weighting. A theoretical formulation was completed, and a list of candidate algorithms developed. Work is continuing on the design of a testing program for comparison of candidates.

ALLFIT is a program developed by DeLean, Munson and Rodbard at NIH to permit approximation by mathematical formulas (curve-fitting) of data from related families of experiments. It has many users within the NIH intramural community and elsewhere. ALLFIT allows variations of a single type of mathematical model (called logistic). The commercial software package Mathematica™ has an advanced programming language with capabilities for description and manipulation of symbolic mathematical formulas. An extension of ALLFIT using the Mathematica™ programming language has been designed and is under development. It should permit curve-fitting of related data using many different models, which can be described by user specification of Mathematica™ expressions. This has the additional advantage of making extended ALLFIT capabilities available simultaneously for DOS/Windows™ and Macintosh® microcomputers, UNIX® workstations and the Convex supercomputer at NIH.

Impact/Value

Neural network training with unequal error weighting addresses some of the problems with development of diagnostic screening tests. Maximization of screening cost-effectiveness translates naturally into optimization of test outcome decision algorithms with unequal error penalties.

The reformulation of ALLFIT using Mathematica™ may lead to a more flexible and widely available tool for analysis of related families of data, and provide a model for exploitation of advanced language capabilities. ALLFIT is already widely used for analysis of dose-response curves from bioassays, radioreceptor assays, radioimmunoassays and DNA-RNA hybridization.

Proposed Course

Candidate methods for training neural networks with unequal error weighting will be tested and compared. As needed, further refinements may be introduced to improve performance.

The ALLFIT for Mathematica™ program extension will be completed and tested.

Publications and Presentations

Bezrukov S. M., Kasianowicz J. J. Fluctuations in current through a single open ion channel reveal titration kinetics of ionizable residues, *Phys Rev Lett* 1993; 70:2352-55.

Bezrukov S. M., Vodyanoy I. On noise in biological membranes and relevant ionic systems. Review in "Advances in Chemistry Series No. 235. Membrane Electrochemistry," 1993 (in press).

Bezrukov S. M., Vodyanoy I., Parsegian V. A. Delineation of channel structures by the osmotic

action and penetration of differently sized neutral polymers, *Biophys J* 1993; 64:A92.

Bezrukov S. M., Vodyanoy I. Probing alamethicin channels with water-soluble polymers. Effect on conductance of channel states, *Biophys J* 1993; 64:16-25.

Bloor J. E., Eckert-Maksic M., Hodoscek M., Maksic Z. B., Poljanec K. *Ab Initio* calculations of the Mills-Nixon effect in Indan, Tetralin, and in related systems, *New J Chem* 1993; 17:157-60.

Brooks B. R., Hodoscek M. Parallelization of CHARMM for MIMD machines. In: Massively Parallel Processing Supercomputing Series, Chemical Design Automation News 1992; 7(12):16-22.

Cohen J. A., Parsegian V. A., Rau D. C. Osmotic pressure of 3-dimensional ordered colloidal suspensions, *Biophys J* 1993; 64:A63.

Durand D., Field M. J., Quilichini M., Smith J. C. Lattice vibrations in crystalline L-Alanine, *Biopolymers* 1993.

Eckert-Maksic M., Hodoscek M., Maksic Z. B., Poljanec K. Mills-Nixon effect in heteroanalogues of cyclopropanbenzene, *Int J Quant Chem* 1992; 42: 869-77.

Eckert-Maksic M., Maksimovic L., Hodoscek M. Electronic structure of fused 7-oxanorbornenes. photoelectronspectroscopic study, *Tetr Lett* 1993; 34:4245.

Eckert-Maksic M., Maksic Z. B., Hodoscek M., Kovacek D., Rupnik K. Intra- and extra-molecular electrostatic potentials in vitamin C, *J Mol Struct (Theochem)* 1992; 256:271-86.

Fang Y., Rand R. P., Leikin S., Kozlov M. M. Chain-melting reentrant transition in bimolecular layers at large separation, *Phys Rev Lett* 1993; 70: 3623-26.

Gawrisch K., Ruston D., Zimmerberg J., Parsegian V. A., Rand R. P., Fuller N. Membrane dipole potentials, hydration forces, and the ordering of water at membrane surfaces, *Biophys J* 1992; 61:1213-23.

Hadzi D., Hodoscek M., Grdadolnik J., Avbelj F. Intermolecular effects on phosphate frequencies in phospholipids – infrared study and *ab initio* model calculation, *J Mol Struct (Theochem)* 1992; 266:9-19.

Hodoscek M., Kovacek D., Maksic Z. B. Theoretical study of Mills-Nixon effect in naphto-cyclobutenes and -cyclobutadienes, *Theor Chim Acta* (Berlin) 1993; 86(4):343-51.

Hodoscek M., Kovacek D., Maksic Z. B. Influence of substituents on the Mills-Nixon effect in some naphtodicyclobutenes and naphtodicyclobutadienes, *J Mol Struct (Theochem)* 1993; 100:213-20.

Hutchinson, G. Relation categories and coproduct congruence categories in universal algebra, *Algebra Universalis* 1993 (in press).

Keller S. L., Bezrukov S. M., Gruner S. M., Tate M. W., Vodyanoy I., Parsegian V. A. Probability of alamethicin conductance states correlates with non-lamellar tendency of bilayer phospholipids, *Biophys J* 1993; 65:23-7.

Kornyshev A. A., Kosskowski D. A., Leikin S. Surface phase transitions and hydration forces, *J Chem Phys* 1992; 97:6809-19.

Kornyshev A. A., Leikin S. Theory of hydration forces. In: Lipowsky R., Richter D., Kremer K., eds: The Structure and Conformation of Amphiphilic Membranes. Springer-Verlag: Berlin 1992; 66:83-6.

Leikin S. Hydration forces in protein-protein recognition, *Bull Amer Physical Soc* 1993; 38:495.

Leikin S., Rau D. C., Parsegian V. A. Direct measurement of forces between self-assembled proteins: temperature-dependent exponential forces between collagen triple helices, *Proc Nat Acad Sci* 1993 (in press).

Leikin S., Rau D. C., Parsegian V. A. Temperature-dependent forces measured between collagen triple helices, *Biophys J* 1993; 64:A270.

Leikin S., Parsegian V. A., Rau D. C., Rand R. P. Hydration forces, *Ann Rev Phys Chem* 1993; 44:369-95.

McKinnon S. J., Whittenburg S. L., Brooks B. R. Molecular dynamics simulation of oxygen diffusion through hexadecane monolayers with varying concentrations of cholesterol, *J Phys Chem* 1992; 96:10497-506.

Milne G. W. A., Nicklaus M., Hodoscek M. Molecular modeling in solvent, *J Mol Struct (Theochem)* 1993; 291(1):89-103.

Munson, P. J. Data visualization techniques for protein structure, NIH Research Festival, Bethesda, MD September 1993 (poster).

Munson, P. J. Pattern recognition of protein structures and substructures, NIH Research Festival, Bethesda, MD September 1993 (poster).

Munson, P. J. Secondary structure prediction using penalized likelihood, Computer Science and Statistics Interface, San Diego, CA April 1993 (presentation).

Munson, P. J. Semiparametric and kernel density estimation procedures for prediction of protein secondary structure, ASA Annual Meeting, San Francisco, CA August 1993 (presentation).

Munson, P. J. Semiparametric procedures for protein secondary structure prediction, Mt. Sinai School of Medicine, New York, NY October 1992 (presentation).

Munson, P. J. Semiparametric statistical methods for protein secondary structure prediction, Biomedical Simulation Resource Workshop, Los Angeles, CA May 1993 (presentation).

Munson, P. J. Statistical methods for protein secondary structure prediction, Intelligent Systems in Molecular Biology, Bethesda, MD July 1993 (poster).

Munson P. J., Cao L., Di Francesco V., Porrelli R. Semiparametric and kernel density estimation procedures for prediction of protein secondary structure. American Statistical Association Annual Meeting, Statistical Computing Section, 1993 (in press).

Munson, P. J, Taylor R. C., Michaels G. S. DNA Correlations (Scientific Correspondence). *Nature* 1992; 360:636.

Oerter K. E., Kamp G. A., Munson P. J., Nienhuis A. W., Cassorla F. G., Manasco P. K. Multiple hormone deficiences in children with hemochromatosis, *J Clin Endo Metab* 1993; 76(2):357-61.

Osawa Y., Darbyshire J. F., Steinbach P. J., Brooks B. R. Metabolism-based transformation of myoglobin to an oxidase by $BrCCl_3$ and molecular modeling of the oxidase form, *J Bio Chem* 1993; 268:2953-59.

Parsegian V. A., Rand R. P., Rau D. C. Swelling from the perspective of molecular assemblies and single functioning biomolecules, NATO ASI Series H 64, 1992; 623-47.

Parsegian V. A., Zimmerberg J. Channels under osmotic stress. In: Jackson M. B., ed. Thermodynamics of Membrane Receptors and Channels. Boca Raton, Florida, CRC Press, 1993; 389-405.

Parsegian V. A., Gershfeld N. L. Inert glue in the surface force apparatus? Where are the controls? *Biophys J* 1993; 64:A222.

Poljanec K., Hodoscek M., Kobal I. *Ab initio* calculations of stationary points on the potential energy surface and determination of kinetic isotope effects for the reaction of CO with Cu_2O,In: Cluster Models for Surface and Bulk Phenomena, G. Pacchioni, *et al*, eds., Plenum Press, New York, 1992.

Porrelli R. N., Munson P. J., Rodbard D. A model for the effect of estrogen antagonists on cooperative estradiol binding, *J Recept Res* 1993; 13(7):1055-81.

Steinbach P. J., Brooks B. R. Protein hydration elucidated by molecular dynamics simulation, *Proc Natl Acad Sci* 1993 (in press).

Tsao Y-H., Evans D. F., Rand R. P., Parsegian V. A. Osmotic stress measurements of dihexadecylmethylammonium acetate bilayers as a function of temperature and added salt, *Langmuir* 1993; 9:233-41.

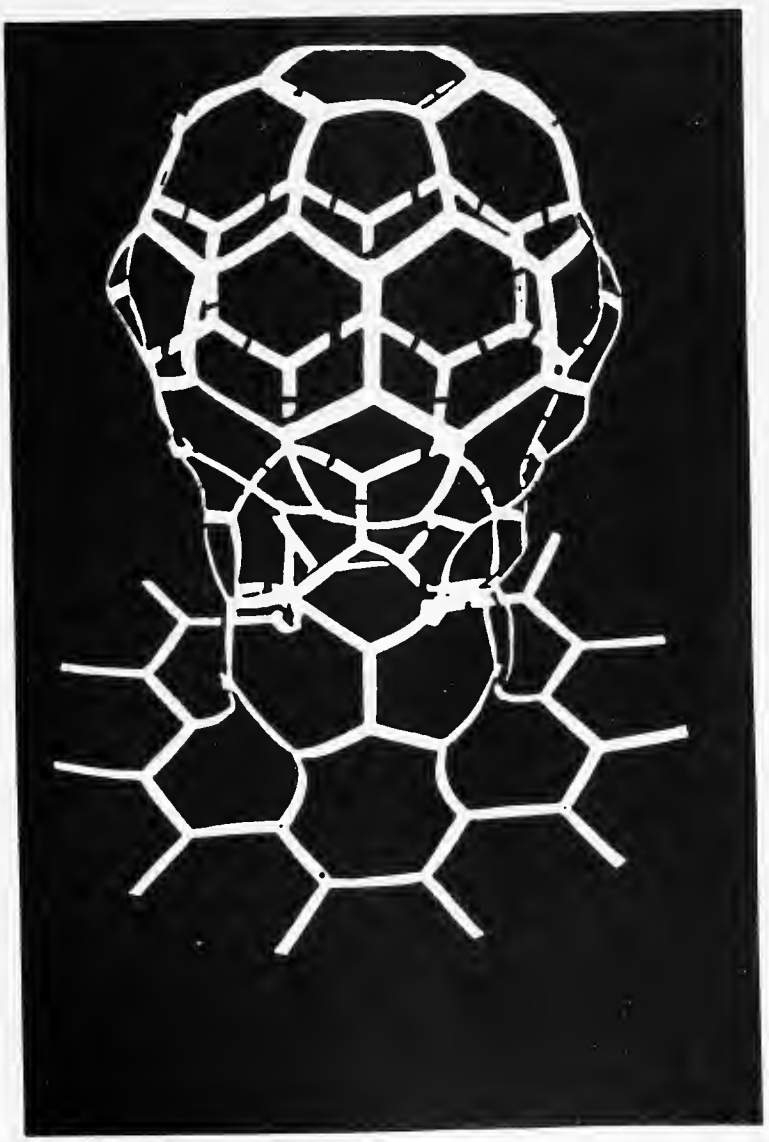
Venable R. M., Brooks B. R., Carson F. W. Theoretical studies of relaxation of a monomeric subunit of Human Immunodeficiency Virus Type 1 protease in water using molecular dynamics, *Proteins Struct Func. and Gen* 1993; 15:374-84.

Vikic-Topic D., Hodoscek M., Graovac A., Becker E. D., Lodder G., Zuilhof H. On the calculations of deuterium long range isotope effects on carbon-13 chemical shifts. In: Nuclear Magnetic Shieldings and Molecular Structure, J. Tossell, ed., NATO ASI Series C, Vol 386, Kluwer Academic Publishers, Dordrecht 1993, p. 574.

Vodyanoy I., Bezrukov S. M. Sizing of an ion pore by access resistance measurements, *Biophys J* 1992; 62:10-11.

PSL

Physical Sciences Laboratory



Physical Sciences Laboratory

George Weiss, Ph.D., Chief

Members of the Physical Sciences Laboratory (PSL) develop and apply techniques of experimental and theoretical physics and applied mathematics related to the biomedical sciences. The work of the PSL consists both of original research projects and of consulting to NIH investigators in areas of the laboratory's expertise. Dr. Nossal leads a group which works in biological application of optical methods and neutron scattering techniques. This project requires both laboratory and theoretical work. Dr. Nossal's group shares laboratory space with Biomedical Engineering and Instrumentation Program/NCRR for the optical component of their work, and the neutron scattering work is done using laboratory facilities at the National Institute of Standards and Technology.

An important component of the PSL is that of consulting with NIH investigators using sophisticated mathematical tools. This includes methods for the solution of nonlinear equations which find wide application in biochemical investigations, the development of optimization techniques as applied to the design of biochemical experiments, and the development of mathematical models in collaboration with a number of experimental biologists related to their work. This aspect of the PSL's work is implemented by Richard Shrager and George Weiss.

Research Projects

Biophysical Analysis

R. J. Nossal, Ph.D.
with A. Gandjbakhche, Ph.D., A. J. Jin, Ph.D., G. H. Weiss, Ph.D. (DCRT/PSL); R. Bonner, Ph.D., J. Schmitt, Ph.D. (NCRR/BEIP); J. C. Calvo, Ph.D., V. Hascall, Ph.D., M. Yanagishita, M.D. (NIDR/BRB); E. Kocsis, Ph. D., A. C. Steven, Ph.D. (NIAMS/LSB); A. P. Andrews, Ph.D., S. Krueger, Ph.D. (NIST); R. Agah, M.D. (Methodist Hospital, Houston); M. Motamedi, Ph.D. (University of Texas Medical Center); S. Havlin, Ph.D. (Bar-Ilan University); P. Mills, Ph.D. (Universite de Paris)

Quantitative physical and mathematical methods have been applied to several research problems of potentially broad interest to biomedical scientists. These include: 1) the use of mathematical and physical techniques to understand the properties and function of biological materials as they relate to cell function. Emphasis has been one of the priorities of polymeric assemblies such as those found on, or near, membrane surfaces or lying between cells. 2) the development of new scattering techniques – using neutrons or light – to examine biological macromolecules when present in concentrated solutions or within highly complex structures, and 3) the analysis of schemes which utilize light to probe the physiological status of biological tissues. Many of these studies are interrelated; in some cases our emphasis is on developing new physical methods; others are directed towards gaining broad insight into aspects of cell function.

Cell Biophysics

Clathrin is a protein which is widely distributed in eukaryotic cells. A dynamic cycle of clathrin lattice assembly and disassembly is a critical element of receptor-mediated endocytosis. The latter step is an important mechanism by which the cells take up materials from extracellular fluid. We have studied various aspects of the rearrangements taking place when clathrin-coated pits, found on the surface of the cells, transform into the coated vesicles which carry accumulated receptor-ligand complexes into the cell. By invoking some basic universal topological rules, we have uncovered stringent constraints which govern these transformations. We were able to infer logical mechanisms for the budding of clathrin-coated vesicles from the cell surface. We found that complexes of clathrin molecules (known as triskelions) must add by pairs at any basic transformation step, and that additions of triskelions to the interiors of coated pits are the

most likely pathway for lattice transformation to occur. (Earlier models postulated a propagation of lattice rearrangements from the edge of the pits.) A collaborative study has been initiated with other NIH scientists to evaluate the energetic and statistical factors involved in coated pit transformation. In this collaboration, we aim to understand how the binding of ligands to pit-associated receptors might affect the endocytotic cycle.

Scattering Techniques

Small-angle neutron scattering facilities at the National Institutes of Standards and Technology research reactor have been used to acquire data on polymer gels. Attention recently has been given to studying the structure of agarose, which has long been familiar to biochemists and molecular biologists as a matrix for electrophoretic separation of macromolecules. Characteristics such as void size, strand thickness, and gel microheterogeneity – as well as changes that might occur when electric fields are applied – have been investigated. This is the initial stage of a study which, in part, has been undertaken to develop methods for examining biological materials such as actin gels and proteoglycan matrix material. Information obtained for agarose gels will be useful in understanding how gel microstructure depends on factors such as polymer concentration and solvent quality. In a companion study, mathematical expressions have been derived to explain laser quasielastic light scattering from particles moving through disordered, multiply-scattering immobile structures. One can relate the time dependence of the photon autocorrelation function to the varying structure of the matrix. Potential applications concern determinations of blood flow within bones.

Tissue Optics

Emphasis has been on providing a theoretical basis for quantitative use of light for medical diagnosis and therapy. Recent work has focused on understanding how light passes through tissue of defined thickness, and how the intensity distributions of light discerned in transillumination measurements depend on tissue optical parameters. Computer simulations were performed to substantiate derived mathematical expressions for photon pathlength distributions and surface intensities. The resulting theory has been applied to the analysis of data acquired in a collaborative study concerning the energetics of thermal damage to tissue. This study has as its ultimate goal a deeper understanding of how laser light affects tissue during surgery. Another application has been to discover the resolution limits for optical transillumination of abnormalities which are deeply embedded in tissue. Resolution is proportional to the square root of the time of observation. However, when factors such as tissue heterogeneity are taken into account, detection of abnormalities becomes very difficult. Results suggest that a time-resolved measurement most likely will be useful when the scattering cross-sections of the normal tissue components are similar.

Instrumentation Analysis

G. H. Weiss, Ph.D.
with R. Shrager, H. Taitelbaum, Ph.D. (DCRT/PSL); J. A. Ferretti, Ph.D. (NHLBI/IR); R. Spencer, M.D., Ph.D. (NIA/IRP); S. Havlin, Ph.D. (Bar-Ilan University); U. Shmueli, Ph.D. (Tel-Aviv University)

A project of continuing interest in the PSL is that of optimizing nuclear magnetic resonance (NMR) experiments. The past year has seen the development of the theory of two-stage experiments for measurements of spin-lattice relaxation times (T_1). These experiments mimic the way spectroscopists make such measurements, although they are generally carried out without any formal experimental design. The designs developed in the PSL yield the most precise values of T_1 for a given amount of spectrometer running time. Precision of such measurements is a *desideratum* in applications of NMR to both medical imaging and to the determina-

tion of molecular configurations; spectrometer running time is often an important limitation on such measurements. A second related project is that of determining reaction rates *in vivo* in physiological systems using NMR techniques. This type of experiment is known as the saturation transfer experiment. In such experiments it is extremely important to make such measurements as quickly as possible as parameters of the system can change during the course of the measurement. Techniques for optimizing such measurements by reducing, insofar as possible, the time to measure reaction rates to within a specified precision are presently being developed, and preliminary results are available. The coming year should see the completion of this project for the simplest formulation of reaction kinetics. A natural continuation of this project is the optimal design of experiments aimed at measuring rates of more complicated chemical reactions of physiological interest. This will involve a combination of analysis and simulation.

There are a number of techniques for compensating crystallographic data for the effects of noisy data. Such corrections are vital for calculating structural information from scattering experiments. The theory underlying these corrections uses the assumption that the noise has a specific (Poisson) distribution. A systematic experimental study of such noise has been undertaken with Professor Uri Shmueli of Tel-Aviv University. The initial findings suggest that the distribution of noise depends on the scattering angle and can differ from the customarily assumed distribution. Future research in this area will be directed towards an examination of the implications of the experimentally observed differences, with a view towards developing improved methods for compensating for the effects of noise.

Work continues on the development of simplified approximations to the complicated but exact results previously obtained by us for direct phase determination. These approximations have allowed the replacement of very tedious and computer-intensive calculations by formulae that very often can be evaluated without the use of a computer. Much of the work of the past 10 years on exact representations of the functions has been summarized in a nearly completed monograph by Uri Shmueli and George Weiss, *Introduction to Crystallographic Statistics*. This is to be published by Oxford University Press.

The PSL has undertaken to assist the Nuclear Medicine Branch of the Clinical Center in a number of aspects of their research. Initially, this is likely to take the form of the development of a number of simulation packages required for evaluation of a number of schemes for image reconstruction, and the development of mathematical models to aid in the optimization of imaging modalities.

Studies in Applied Mathematics and Statistics

G. H. Weiss, Ph.D.
with S. Havlin, Ph.D., M. Gitterman, Ph.D. (Bar-Ilan University); J. Masoliver (University of Barcelona); R. Kopelman, Ph.D. (University of Michigan); H. Larralde (Boston University); A. Yergey, Ph.D., R. Goans, M.D. (NICHD/LTPB)

This project includes the application of the theory of diffusion and extensions thereof to a number of problems in chemistry and biology. A "singular perturbation theory" has been developed for the solution of reaction-diffusion problems. This theory applies to situations in which the diffusive component is relatively weak in comparison to a convective force. This is commonly the case in electrophoretic and chromatographic systems. In the past year we have compared two different methods that have been proposed for solving such problems, finding that each of them can be used to advantage for the solution of different problems. Approximate solutions have proven useful in characterizing separation properties of chromatographic or electrophoretic systems having nonuniform spatial properties. Preliminary work is in progress on the development of a theory of calcium absorption by bone in various age groups. The model being tested

is very close to chromatographic models developed in the PSL, and the object of doing so is to test the possibility of developing experimentally implementable laboratory models to test a number of hypotheses.

A number of problems related to optical imaging have been studied using techniques developed in random walk theory. One such problem relates to the ability of time-resolved transillumination experiments to detect objects in tissue, e.g., tumors whose optical absorption exceeds that of the surrounding tissue. The random walk theory is useful in simplifying many more cumbersome calculations or simulations that have been used by workers in optical techniques. This allows the examination of changes in the parameters available to the experimenter, and a determination of the possibility of using optical imaging techniques in specific situations. A continuation of this type of analysis to consider problems related to resolution issues is a natural outcome of this work.

A book, *Aspects of the Random Walk*, by George Weiss has been accepted for publication by North-Holland Press and should appear by year's end. A collection of articles, *Contemporary Problems in Statistical Physics*, has been assembled by George Weiss for the Society for Industrial and Applied Mathematics and will be published in their series *Frontiers in Applied Mathematics*.

Mathematical and Computational Methods for Solving Nonlinear Equations

R. I. Shrager
with G. H. Weiss, Ph.D. (DCRT/PSL); S. Bose, Ph.D. (J. Nehru University, New Delhi, India); R. Berger, Ph.D., R. Hendler, Ph.D. (NHLBI/LCB); Z. Dancshazy, Ph.D. (Hungarian Acad. Sci.); M. L. Doyle, Ph.D., D. W. Myers, Ph.D., and G. K. Ackers, Ph.D. (Dept. of Biochem, Washington University School of Medicine, St. Louis); K. D. Vandegriff, Ph.D. (Letterman Army Institute of Research); M. Perrella, Ph.D. (University of Milan, Italy); R. Carson, Ph.D. (CC/NMD); U. Shmueli, Ph.D. (Tel-Aviv University, Israel)

The purpose of this project is to provide NIH investigators with mathematical tools for insight, analysis, and solution of complex equations arising in the modeling of biological systems. To facilitate these efforts, PSL develops mathematical methods that are accessible to investigators from many disciplines. Software packages based on these developments are made available to the research community as general research tools. Advice on the use of certain commercial mathematical software packages is also offered.

• Binding Rates of Hemoglobin (Hb) to Various Ligands (with M. Perrella, Ph.D. University of Milan). An algorithm has been developed for computing the time course of Hb-ligand binding which allows the curve fitting of binding rates. The use of this algorithm allows savings in computer time of orders of magnitude in comparison to previously applied techniques.

• Conformational Changes in Hemoglobin (Hb) Binding (with K. Vandegriff, Ph.D., Letterman Army Institute of Research). Hb-oxygen binding is studied using singular value decomposition (SVD) and the most precise optical spectra that currently available equipment can provide, in an effort to detect conformational changes.

• Relaxation Kinetics of Bacteriorhodopsin (bR) (with R. Hendler, Ph.D., NHLBI/LCB; Z. Danceshazy, Ph.D., Hungarian Acad. Sci.; S. Bose, Ph.D., J. Nehru University, New Delhi, India). The kinetics of bR after laser flash seem to depend on the intensity of the flash. Explanations for this dependence are being sought using SVD in conjunction with target theory. Our work has shown that both cooperative and noncooperative models are capable of mimicking the features of the relaxation.

• Regression Analysis of Oxygenation Isotherms (with M. Doyle, Ph.D., D. Myers, Ph.D., and G. Ackers, Ph.D., Washington University School of Medicine, St. Louis). Extensive simulations have been run using different assumptions about

experimental error. Methods have been devised for generating first estimates of parameters and for solving total least squares equations efficiently.

• Concentrations of ADP and ATP by Partial Least Squares (PLS) Methods (with R. Berger, Ph.D., NHLBI/LCB). This project is concerned with the uses and pitfalls of PLS in chemical analysis, especially in trying to find small signals in the presence of noise. Several computer codes have been written to implement the use of PLS in this project.

• Rapid Computation of the Probability Density Function (pdf) for the Three-Phase Invariant Used in Direct Methods of Phase Determination in X-Ray Crystallography (with U. Shmueli, Ph.D., Tel-Aviv University, Israel; G. Weiss, Ph.D., DCRT/PSL). Exact expressions are available for this frequently used function in programs for reducing crystallographic data, but these require daunting numerical calculations to obtain usable results. It is possible to develop simpler, but approximate, expressions from the more accurate calculations. This program is presently being implemented and should be completed shortly.

• Imaging Regional Cerebral Blood Flow (with R. Carson, Ph.D., CC/Nuclear Medicine Dept.). A proposed method for computing this quantity without an explicit measurement of the associated arterial blood flow has been improved upon at PSL. The theory for doing this is in hand but further testing on actual data is required. These data are presently being assembled in the Nuclear Medicine Department to test the utility of the proposed method.

• Kinetics of Reduction of Cytochrome aa3 (with R. Hendler, Ph.D., CC/NMD; S. Bose, Ph.D., J. Nehru University, New Delhi, India). The dynamics of cytochrome reduction are being observed with a rapid-scan multiwavelength spectrophotometer. Its output is analyzed in terms of singular value decomposition as well as other computational methods. This project is a continuing one since experimental data continue to be collected.

Publications

Ben-Naim E., Redner S., Weiss G. Partial absorption and "virtual traps," *J Stat Phys* 1993; 71:75-88.

Calvo J. C., Gandjbakhche A. H., Nossal R., Hascall V., Yanagashita M. Rheological effects of the presence of hyaluronic acid in the extracellular media of differentiated 3T3-L1 preadipocyte cultures, *Arch Biochem Biophys* 1993; 302:475-86.

Colombo M. F., Rau D. C., Parsegian V. A. The role of water in hemoglobin function and stability-response, *Science* 1993; 259:1336.

Dayan I., Havlin S., Weiss G. H. Laser beam spreading in transmission through a slab, Lasers in the Life Sciences 1993 (in press).

Doueck P., Gandjbakhche A. H., Leon M. B., Bonner R. F. Functional properties of a new rheolytic catheter for percutaneous thrombectomy: *in vitro* investigation, *J Invest Radiol* 1993 (in press).

Doyle M. L., Myers D. W., Ackers G. K., Shrager R. I. Weighted nonlinear regression analysis of oxygenation isotherms, In: Everse J., Winslow R. M., Vandegriff K. D., eds. *Methods in Enzymology: Hemoglobin*, Orlando, FL, Academic Press 1993 (in press).

Eisenberg E., Havlin S., Weiss G. H. Diffusive fluctuations in different realizations of a random medium, *Phys Rev E* 1993 (in press).

Gandjbakhche A. H., Nossal R., Bonner R. F. Scaling relationships for theories of anisotropic random walks applied to tissue optics, *Appl Opt* 1993; 32:504-16.

Gandjbakhche A. H., Weiss G. H., Bonner R. F., Nossal R. Photon pathlength distributions for transmission through optically turbid slabs, *Phys Rev E* 1993; 48:810-18.

Gandjbakhche A. H., Taitelbaum H., Weiss G. H. Random walk analysis of time-resolved transillumination measurements in optical imaging, *Physica A* 1993; 200:212-21.

Gandjbakhche A. H., Schmitt J. M., Bonner R., Nossal R. Random walk theory applied to noninvasive *in vivo* optical measurements of human tissue. In: Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Paris: Inst Elect Electron Engr, Piscataway, NJ, 1992; 332-333.

Gitterman M., Weiss G. H.. A comment on early-time solutions of the Smoluchowski equation, *J Stat Phys* 1993 (in press).

Gitterman M., Weiss G. H. A comparison of two methods for solving transport equations with weak diffusion, *Sep Sci Tech* 1993 (in press).

Gitterman M, Weiss G. H. A singular perturbation theory for reaction diffusion equations, *Chem Phys* 1993 (in press).

Gitterman M., Weiss G. H. A transition in a noisy linear system driven by a periodic signal, *J Stat Phys* 1993 (in press).

Gitterman M., Weiss G. H. "Escape" of a periodically driven particle from a metastable state in a noisy system, *J Stat Phys* 1993; 70:107-23.

Gitterman M., Weiss G. H. Generalized theory of the kinetics of tracers in biological systems, *J Math Biol* 1993 (in press).

Gitterman M., Weiss G. H. Small-noise approximations to the solution of the Smoluchowski equation, *Phys Rev E* 1993; 47:976-80.

Havlin S., Kiefer J. E., Trus B., Weiss G. H., Nossal R. On the detectability of inclusions hidden in optically turbid tissue, *Appl Opt* 1993; 32:617-27.

Havlin S., Kiefer J. E., Trus B., Weiss G. H., Nossal R. Numerical method for studying the detectability of inclusions hidden in optically turbid tissue, *Appl Opt* 1993; 32:617-27.

Hemric M. E., Lu F., Shrager R., Carey J., Chalovich J. M. Reversal of caldesmon binding to myosin with calcium-calmodulin or by phosphorylating caldesmon, *J Biol Chem* 1993; 20:15305-11.

Hendler R., Bose S., and Shrager R. Multiware analysis of the kinetics of reduction of cytochrome aa3 by cytochrome c, *Biophys J* (in press).

Hendler R. W., Dancshazy Z., Bose S., Shrager R. I., Tokaji Z. Influence of excitation energy on the bacteriorrhodopsin photocycle, *Biophys J* 1993 (in press).

Hendler R., Shrager R. I. Deconvolutions based on singular value decomposition and the pseudoinverse. A guide for beginners, *J Biochem Biophys Meth* 1993 (in press).

Jin A. J., Fisher M. E. Effective interface Hamiltonians for short-range critical wetting, *Phys Rev B* 1993; 47:7365-88.

Jin A. J., Fisher M. E. Stiffness instability in short-range critical wetting, *Phys Rev B* 1993; 48:2642-58.

Jin A. J., Nossal R. Topological mechanisms involved in the formation of clathrin-coated vesicles, *Biophys J* 1993; 65:1523-37.

Krueger S., Andrews A. P., Nossal, R. Small angle neutron scattering studies of structural characteristics of agarose gels, *Biophys Chem* 1994 (in press).

Masoliver J., Porra J. M., Weiss G. H. Some two and three-dimensional persistent random walks, *Physica A* 1993; 193:469-82.

Masoliver J., Weiss G. H. On the maximum displacement of a one-dimensional diffusion process

described by the telegrapher's equation, *Physica A* 1993; 195:93-100.

Masoliver J., Porra J. M., Weiss G. H. Solution to the telegrapher's equation in the presence of reflecting and partly reflecting boundaries, *Phys Rev E* 1993 (in press).

Nossal, R. Analysis of laser Doppler measurements of blood flow in statistically irregular media, In: Nossal R., Pecora R., Priezzhev A. V., eds. Society of Photo-Instr Engr (SPIE) Proceedings, Bellingham, WA, 1993; 1884:118-24.

Perrella M., Shrager R. I., Ripamonti M., Manfredi G., Berger R. L., Rossi-Bernardi L. Mechanism of the oxidation of deoxyhemoglobin as studied by isolation of the intermediates suggests tertiary structure dependent cooperativity, *Biochem* 1993; 32(19):5233-38.

Posner Y., Shmueli U., Weiss G. H. Exact conditional distribution of a three-phase invariant in the space group P1. III. Construction of an improved Cochran-like approximation, *Acta Cryst A* 1993; 49:260-265.

Schach R., Shmueli U., Goldberg I. Some statistics of background radiation from observed diffraction profiles, *Acta Cryst A* 1993 (in press).

Schmitt J. M., Knuttel A., Gandjbakhche A. H., Bonner R. F. Optical characterization of dense tissue using low-coherence interferometry, In: Society of Photo-Instr Engr (SPIE) Proceedings, Bellingham, WA, 1993; 1889:197-210.

Shmueli U., Stein Z., Weiss G. H. Developments in the study of effects of space-group symmetry and atomic heterogeneity on intensity statistics, *Acta Chim Hung* 1993; 130:261-78.

Shmueli U., Weiss G. H. Effects of non-crystallographic symmetry on the E_1 relationship. I. Bicentric arrangements in P1, *Acta Cryst* 1993 (in press).

Shmueli U., Weiss G. H. *Introduction to Crystallographic Statistics*, Oxford Univ Press 1993 (in press).

Shrager R. Analytic models for nonlinear curve-fitting of forward-rate binding data, with applications to hemoglobin, *J Biochem Biophys Meth* 1992; 25:113-24.

Shrager R. I. Modeling chemical reaction: the Jacobian paradigm and related issues. *Methods in Enzymology: Numerical Methods* Orlando, FL, Academic Press 1993 (in press).

Shrager R. I., Shmueli U., Weiss G. H. Exact conditional distribution of a three-phase invariant in space group P1. IV. Further improvements of Cochran-like approximations, *Acta Cryst* 1993 (in press).

Sparling L. C., Weiss G. H. Some effects of beam thickness on photon migration in a turbid medium, *J Mod Opt* 1993; 40:841-59.

Spencer R., Ferretti J. A., Weiss G. H. Spillover and incomplete saturation in kinetic measurements, *J Mag Res* 1993 (in press).

Taitelbaum H. Segregation in reaction-diffusion systems, *Physica A* 1993 (in press).

Taitelbaum H., Ferretti J. A., Spencer R. G. S., Weiss G. H. Optimization of two-stage measurements of T_1, *J Mag Res* 1993 (in press).

Taitelbaum H., Ferretti J. A., Spencer R. G. S., Weiss G. H. Two-stage inversion recovery experiments for measurement of T_1, *J Mag Res* 1993 (in press).

Taitelbaum H., Weiss G. H. Segregation at a single-trap in the presence of fields, *Mat Res Soc Symp* 1993; 290:351-360.

Tsao Y. H., Evans D. F., Rand R. P., Parsegian V. A. Osmotic stress measurements of dihexadecyldimethylammonium acetate bilayers as a function of temperature and added salt, *Langmuir* 1993; 9:233-41.

Vandegriff K. D., Shrager R. I. Evaluation of oxygen binding to hemoglobin by rapid-scanning spectrophotometry and singular value decomposition. In: Everse J., Winslow R. M., eds. *Methods in Enzymology: Hemoglobin*, Orlando, Florida, Academic Press 1993 (in press).

Vandegriff K. D., Shrager R. I. Pseudo-Equilibrium Studies of Oxygen Binding. In: Everse J., Winslow R. M., eds. *Methods in Enzymology: Hemoglobin*, Orlando, Florida, Academic Press 1993 (in press).

Weiss G. H. A primer of random walkology. In: Havlin S., Bunde A., eds., *Fractals and Disordered Systems*, 1993 (in press).

Weiss G. H. *Aspects of the Random Walk*, North Holland Press 1993 (in press).

Weiss G. H. Contemporary problems in statistical physics, In: *Frontiers in Applied Mathematics*, Soc Ind Appl Math 1993 (in press).

Weiss G. H. Nearest-neighbor distance to a trap in a one-dimensional Smoluchowski model, *Physica A* 1993; 192:617-27.

Weiss G. H., Dishon M., Long A. M., Bendler J. J., Jones A. A., Inglefield P. T., Bandis A. NMR relaxation in polymers using the Kohlrausch-Williams/Watts (KWW) decay function, *Polymer* 1993 (in press).

Weiss G. H., Gitterman M. Motion in a periodic potential driven by rectangular pulses, *J Stat Phys* 1993; 70:93-105.

Weiss G. H., Shmueli U. *Introduction to Crystallographic Statistics*, Oxford University Press, Oxford, England, 1993 (in press).

OAD, OCB

Office of the Associate Director, OCB

Original data from GCG program

```
C       DSNATNSNLERVEYLFLIIFTVEAFLKVIAYGLLFHPNAYLR
        ||| || |||:||| |||||||| |||:|||||| |||:|:|
BCCI    DSNSTNHNLEKVEYAFLIIFTVETFLKIIAYGLLLHPNGYVR
        | || |  |||:|| || :|::|  :|||||| | |  :|:|
Skm     DNNSLNLGLEKLEYFFLTVFSIEAAMKIIAYGFLFHQDAYLR
```

Highlighted for analysis

```
C       DSNATNSNLERVEYLFLIIFTVEAFLKVIAYGLLFHPNAYLR
        ||| || |||:||| |||||||| |||:|||||| |||:|:|
BCCI    DSNSTNHNLEKVEYAFLIIFTVETFLKIIAYGLLLHPNGYVR
        | || |  |||:|| || :|::|  :|||||| | |  :|:|
Skm     DNNSLNLGLEKLEYFFLTVFSIEAAMKIIAYGFLFHQDAYLR
```

Alignment against |

```
C       DSNATNSNLERVEYLFLIIFTVEAFLKVIAYGLLFHPNAYLR
BCCI    DSNSTNHNLEKVEYAFLIIFTVETFLKIIAYGLLLHPNGYVR
Skm     DNNSLNLGLEKLEYFFLTVFSIEAAMKIIAYGFLFHQDAYLR
```

Reverse font highlighting an alignment to |

```
C       DSNATNSNLERVEYLFLIIFTVEAFLKVIAYGLLFHPNAYLR
BCCI    DSNSTNHNLEKVEYAFLIIFTVETFLKIIAYGLLLHPNGYVR
Skm     DNNSLNLGLEKLEYFFLTVFSIEAAMKIIAYGFLFHQDAYLR
```

Office of the Associate Director, OCB

David Rodbard, M.D., Acting Associate Director

In the DCRT reorganization, the research projects that follow were transferred to report to the OAD, OCB.

The DCRT Image Technology Program, under the leadership of the CC's Dr. Stephen Bacharach, has involved division staff in four separate undertakings:

- 3D alignment of PET transmission scans by maximization of 3D pixel-to-pixel correlation (see p. 105)
- automatic tracking of MRI "Tag" grids (see p. 106)
- maximum likelihood estimation of regional radioactivity concentration (see p. 28)
- computer-guided surgery in Von Hippel Lindau disease (see p. 106).

The Acting Associate Director has maintained several research interests. These include studies of the hormonal responsiveness of preadipocytes (3T3-L1 cells in culture) to progestins and mineralcorticoids. These studies were begun by the director in the Laboratory of Theoretical and Physical Biology, NICHD with Dr. C. Rondinone (now NIDDK) and Dr. Michael Baker (now University of California, San Diego). Several studies of mathematical modeling of signal transduction for hormone receptor systems have been completed. These include studies of G-protein mediated membrane receptors and steroid hormone receptors in the nucleus. Other interests include mathematical modeling, curve-fitting and parameter estimation for ligand binding systems and for dose-response curves in general. Several of these studies were undertaken in collaboration with LSB's Dr. Peter Munson, and have led to the development of computer programs that have been disseminated by the thousands worldwide. In the future, previous studies of pulse- and peak-detection for electrophoresis and chromatography will be applied to the challenge of increasing the accuracy and estimating the reliability of base-calling for automated DNA sequence analysis.

Research Projects

DNAdraw for the Macintosh®/Computer Software for DNA Sequence Display and Analysis

M. Shapiro

A previously developed, in-house DOS program for drawing DNA sequences has received considerable use both within the NIH community and more widely. However, based on its inadequacies and numerous requests for a version running on the Macintosh®, work has begun on a completely new version of DNAdraw for the Macintosh®. It will do essentially the same job, i.e., formatting sequence data and drawing highlighted sequences for publication, but it will have a number of significant improvements over the PC version. First, being on the Macintosh® and conforming to the standard Macintosh® principles, it will be immediately usable with little or no reference to a manual. The Macintosh® system of menus will make the specification and drawing of highlights extremely simple for the user. In addition, it will use the capabilities of the mouse to make interaction with the program much easier than the PC version.

Work on the program DNAdraw, written for the Macintosh® computer, has proceeded to the point where Version 1.0 is soon to be released. DNAdraw is based on similar programs written for the DEC®-10, Convex, and PC computers. The use of the Macintosh® user interface makes this new version of DNAdraw much easier to use than previously, and the addition of a number of new features has greatly enhanced the program. In addition to providing the standard sequence highlighting capabilities of changing fonts, shading, proportional spacing, changing styles, underlining, etc., the program has features for automatically highlighting aligned sequence data, and for formatting and translating raw

sequence data. PostScript® output provides the user with publication-quality output. A number of users have tried beta versions of the program, and it should formally be released soon.

Investigation of PCR Primers

M. Shapiro

The Polymerase Chain Reaction (PCR) procedure is a major laboratory technique used in the Human Genome Project for amplification of sequence fragments. One of the difficulties encountered when using PCR is that the product can be contaminated with the DNA of nongenomic sequences, most notably mitochondrial DNA. PCR is initiated by using small sequences, usually 20 or fewer bases, that act as primers for the reaction.

This work is an attempt to recognize primers to avoid, in the sense that they will cause mitochondrial DNA to amplify and contaminate the genomic DNA product. Computer programs were written that, for a given primer, first find all locations of them in human mitochondrial DNA, and then compute whether PCR amplification can be expected, based on combinations of primer locations that are known to cause amplification.

In a collaboration with Dr. Steve Zullo of NIMH, we have developed a program, OLIGFIND, which finds all locations of n-mers (n=6 to 10) in sequence data. The program has been used with mitochondrial sequences to determine the best primers for PCR of genomic sequences. Contamination of genome sequences by mitochondria during PCR is a serious problem, and this work should help prevent such contamination by identifying primer sequences that do not occur in mitochondria, or occur in a fashion that does not interfere with genomic priming.

Computer-Aided Analysis of Electrocardiography

J. J. Bailey, M.D., and E. W. Pottala, Ph.D. with D. Levy, M.D. (NHLBI/Framingham Heart Study), J. E. Norman, Jr., Ph.D. (NHLBI/FSBB), D. MacAreavey (NHLBI/CB); R. W. Bowser (Creighton University), M. Platt (VA Medical Center, Washington, DC)

These studies are directed toward evaluating the prognostic power of the electrocardiogram when analyzed by advanced computer methodology and the predictive accuracy of diagnostic criteria when implemented in ECG computer programs. Appropriate use of digital signal processing in electrocardiography requires application of statistically based techniques of information theory and mathematically based engineering methods as well as knowledge about its clinical relevance.

This project introduced the first computer ECG interpretive program implemented on DCRT's mainframe system, which daily produced interpretations for diagnostic electrocardiograms from the ECG Laboratory in the NIH Clinical Center. Later, project members led a team which supervised the acquisition of the dedicated minicomputer system which has continued to interpret and archive electrocardiographic data for the Clinical Center for a number of years. In the long history of this project, numerous innovative methods for the processing of diagnostic ECGs have been developed and published. A few years ago project members led an international committee in developing standards for digital signal processing in diagnostic ECGs, which were adopted and published by the American Heart Association.

Beginning last year, these studies have been redirected toward the analysis of ambulatory electrocardiography (AECG). Despite extensive literature showing that information extracted by computer analysis of ambulatory electrocardiograms (AECGs) can be related to cardiac risk factors, in this rapidly evolving field there are no standard methods for the routine analysis of AECGs.

Most AECGs are recorded on analog cassette tapes with a slow speed (1 7/8 ips). For that reason a playback device which tracks and corrects for "wow" and other variations in tape speed was necessary for digital data acquisition. Rather than build the

hardware and develop the software to digitize the analog AECGs, DCRT obtained a SpaceLabs FT2000A Medical Analysis and Review Station (workstation), which accomplished this task.

Whether AECG is used in a clinical or research context, the outcome is critically dependent upon the quality and completeness of the data. A particular objective of this research is to carry forward previous work in biosignal analysis (see project report on General Signal Processing for Physiological and Laboratory Data, p. 113) and adapt methodologies to human AECGs with the goal of implementing as much automation as possible to enable and expedite the interpretation of AECG data.

FY93 Progress

In a collaborative study with the Framingham Heart Study and the Field Studies and Biometry Branch, NHLBI, SAS® analysis (using the DCRT mainframe) of fuzzy receiver operating curves (ROCs) showed that adjusting QRS voltage for age, body habitus, and gender significantly improved electrocardiographic criteria for left ventricular hypertrophy, an independent prognostic indicator for fatal cardiac events. Different adjustments in the criteria for women were developed; separate ECG criteria for women which would more accurately predict their risk factors for cardiac events have not, heretofore, been adequately studied.

Last year the SpaceLabs workstation was shown to be compatible with AECG cassette tapes generated by the Scientific Dynacord Model 423 ambulatory ECG recorder used by the NHLBI Clinical Cardiology Branch. This year compatibility with cassettes from a number of different model recorders, including Del Mar Avionics, Marquette, and Cardiodata, all of which are used by the VA Medical Center in Washington, DC, was demonstrated. Tests with the VA Medical Center data revealed some problems with Spacelabs' algorithms for detecting ventricular events. The spectrograms produced by the Spacelabs workstation are primitive and difficult to interpret. Accordingly, the decision was made to use analytical methods with techniques for filtering and sensitive R wave detection previously developed in this laboratory.

Short segments (4 minutes) of AECG data from patients under metronomical control of breathing or under table-tilt manipulations digitized in the Spacelabs workstation were transferred to the laboratory Macintosh® system for analysis; power spectra of RR interval variability demonstrated a feasible method for testing functional autonomic influences upon heart rate.

As it is not feasible to transfer 24 hours of the original AECG data to the laboratory Macintosh® system, it will be necessary to effect a transfer from the SpaceLabs Workstation to the Convex where algorithms for nulliphase filtering, robust R location, and waveform index can reduce the data sufficiently so that the remainder of the analysis can be performed on the Macintosh®. The SpaceLabs workstation and the laboratory Macintosh® system are now connected via TCP/IP which facilitates data transfer to and from the Convex supercomputer.

Future Trends

The tools and algorithms described above should make it possible to pursue clinical studies with 24-hour AECG data from NHLBI, the VA Medical Center, and other possible collaborators. Possible applications to exercise ECGs on subjects being studied in the Laboratory of Cardiovascular Science, NIA have also been discussed.

3D SPECT Image Reconstruction in Nuclear Cardiology

J. J. Bailey, M.D., and E. W. Pottala, Ph.D.
with S. L. Bacharach, Ph.D. (CC/NMD), G. Lan, B.S.
(University of Maryland)

The usual tool for SPECT image reconstruction has been the Fourier transform. The proposed work would investigate the potentiality for more effective filtering and pattern recognition. The literature indicates that the standard filters currently used in 3D image reconstruction may have significant problems with phase distortion. If so, then

there may be considerable advantage in redesigning such filters, even in Fourier space, so as to eliminate phase distortion; this could produce enhancement of edge detection and other kinds of pattern recognition, potentially improving sensitivity and specificity of clinical diagnoses.

FY93 Progress

The Donner Package (FORTRAN II), which does the backprojections and 3D reconstruction, has been transferred from the Nuclear Medicine VAX™ to the Convex. A simple phantom consisting of concentric spheres has been designed, and we set the inner sphere count densities to 0% to 50% of the count densities in the outer sphere.

Future Trends

The simple phantom will be constructed by computer simulation and used in preliminary testing of the standard methods for filtered backprojection. The module which does the filtering will be replaced by newly designed, experimental modules. The resulting slice images can be transferred to a Macintosh® for graphic display.

Nuclear Medicine will collect data from realistic phantoms of normal and abnormal myocardium and transfer the data to the Convex for reconstruction (as above).

The effectiveness of the filter modules will be judged by a number of observers, i.e. their ability to detect known abnormalities in the slice images from a series of experiments as reflected by ROC curves or as demonstrated in two-alternative-forced-choice studies. Further, if the slice images are sectored, then the size of the abnormalities (defects) can be computed and compared with the known, original abnormalities, thereby producing a quantitative analysis.

The results of both qualitative and quantitative analyses will be used to design the optimal filtering scheme, which will take into account: detector energy resolution, count level, pixel size, distance from the collimator face, depth into the scattering media, point or line spread functions, modulation transfer function, anisotropicity of attenuation factors, etc.

Comparing the effectiveness of various filtering schemes, analyzing the effect of correlated vs uncorrelated data from adjacent slices, tailoring filtering parameters to certain given signal or noise characteristics, etc., are the ultimate objectives of this collaboration.

Publications

Bailey J. J., McAreavey D., Pottala E. W. Methods for testing autonomic control of heart rate: table-tilt and metronomic breathing manipulations, *J Electrocard* (Supplement) 1993 (in press).

Costa T., Ogino Y., Munson P. J., Onaran H. O., Rodbard D. Drug efficacy at guanine nucleotide-binding regulatory protein-linked receptors: Thermodynamic interpretation of negative antagonism and of receptor activity in the absence of ligand, *Mol Pharm* 1992; 41:549-60.

Norman J. E. Jr., Levy D., Campbell G., Bailey J. J. Improved detection of echocardiographic left ventricular hypertrophy using a new electrocardiographic algorithm, *J Am Coll Cardiol* 1993; 21:1680-86.

Porrelli R. N., Munson P. J., Rodbard D. A model for the effect of estrogen antagonists on cooperative estradiol binding, *J Recept Res* 1993; 13:1055-81.

Pottala E. W., Bailey J. J., Gilham J. The effect of timing resolution upon RRV spectra with a robust QRS detector after bandpass filtering, *J Electrocard* (Supplement) 1993 (in press).

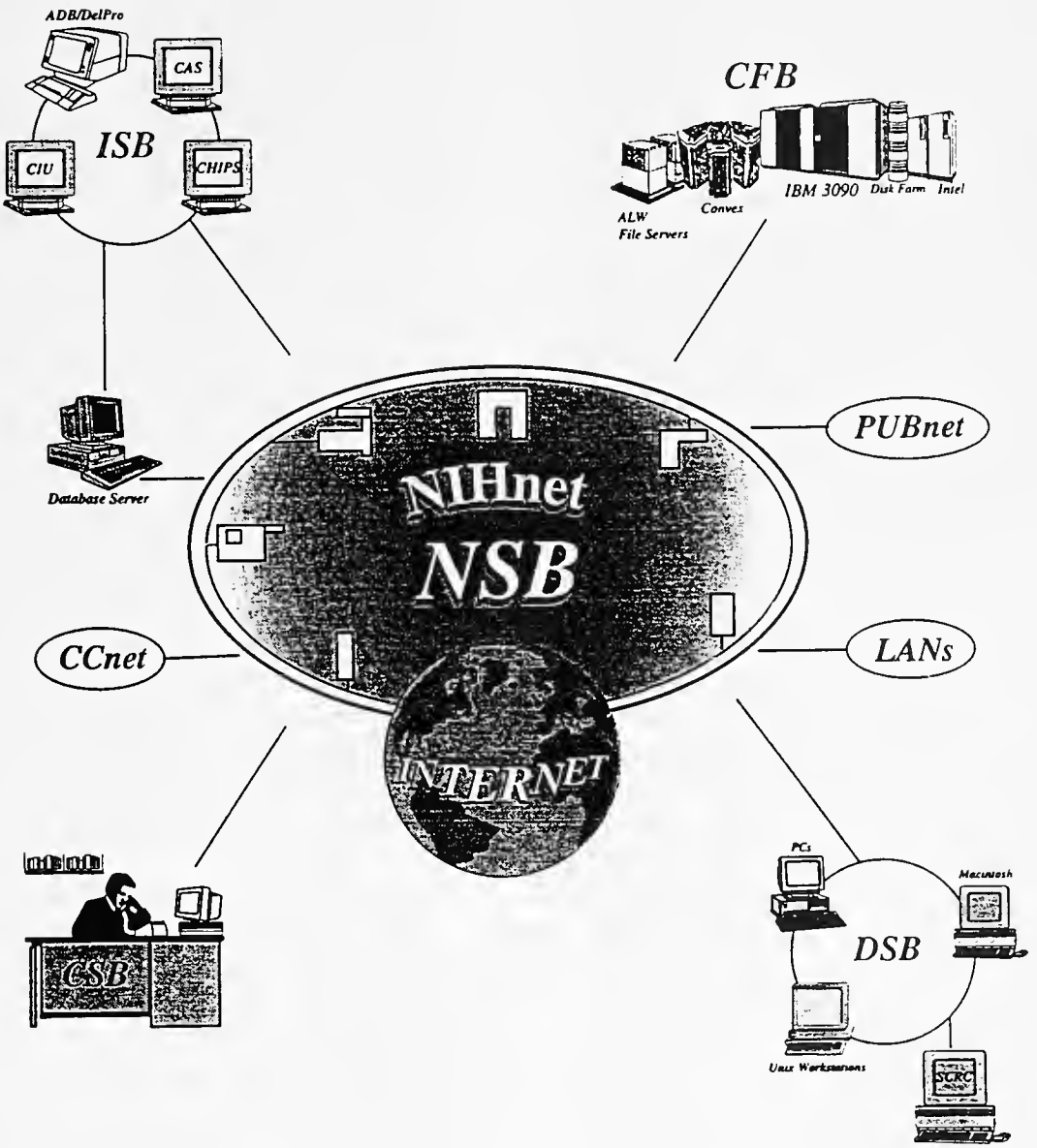
Rondinone C., Baker M., Rodbard D. Progestins stimulate the differentiation of 3T3-L1 preadipocytes, *J Ster Biochem Molec Biol* 1992; 42:795-802.

Rondinone C. M., Baker M., Rodbard D. Aldosterone stimulates differentiation of mouse 3T3-L1 cells into adipocytes, *Endocrin* 1993; 132:2421-26.

Shapiro M. DNAdraw manual, DCRT 1993.

Zullo S., Kennedy J., Gelernter J., Polymeropoulos M., Tallini G., Pakstis A., Shapiro M., Merril C., Kidd K. Eliminating mitochondrial DNA competition for nuclear DNA primers, *PCR Meth Appl* 1993; 3:39-45.

NSB
Network Systems Branch



Network Systems Branch
Harold Ostrow, Chief

Network connectivity has become an essential tool for the biomedical, clinical, and administrative communities at NIH. The NIH-wide area network, known as NIHnet, serves to interconnect local area networks (LANs) throughout NIH – providing the data highway over which images, data, and electronic mail can travel. All of NIH's Institutes, Centers, and Divisions have LANs that depend on NIHnet for wide-area network services. Fiscal Year 1993 saw the creation of the Network Systems Branch (NSB) in DCRT to consolidate support for the several components of the NIHnet networking infrastructure – RESnet, NUnet, CCnet – into a cohesive whole. NIHnet functions and support previously assigned to DCRT's Network Task Group and Computer Center Branch are now handled by the NSB. NSB's mandate includes the following functions:

- design, implement, and monitor the high-speed backbone connections that provide connectivity to local area networks (LANs) on campus
- coordinate, implement, and monitor NIHnet connections to NIH LANs in off-campus buildings and to other networks
- provide guidance and support for locally managed LANs and for "whole building" wiring infrastructure
- design and support DCRT networks, including networks for trans-NIH services and servers
- develop information resources and network-based applications for the NIH community.

The Network Systems Branch is composed of three sections: the *Network Research and Development Section*, the *Customer Support Section*, and the *Systems Development and Support Section*. The branch professional staff includes electronics engineers, computer scientists, and computer specialists. In addition to the efforts of each section, everyone in the branch participates in network monitoring and handles questions and problem reports on the NIHnet Customer Support "hot line."

NIHnet Leverages High-Speed-Fiber and Existing Telecommunications

The NIHnet infrastructure supports over 200 LANs that exist in the Clinical Center, in other on-campus buildings, and in numerous off-campus buildings. This wide range of locations, coupled with the variety inherent in mixed scientific and administrative use, requires flexibility, reliability, and responsiveness on the part of NSB and the NIHnet wiring plant itself. Two connectivity strategies are integrated into NIHnet: a high-speed fiber backbone and a telecommunications-based backbone.

The high-speed fiber backbone, which utilizes the state-of-the-art 100-megabit-per-second FDDI (Fiber Distributed Data Interface) technology, provides the NIHnet connection for over 100 LANs in Buildings 10 (Clinical Center), 12, 12A, 13, 29, 30, 36, 37, and 49. These buildings house a large percentage of the intramural scientists on campus. Their research spans the spectrum of NIH endeavors and can have data-intensive networking requirements. Imaging, full-motion video, genome mapping, and other research applications place heavy demands on the network, with data transmissions that are more time-sensitive and "bursty" than steady. NSB's goal is to provide the NIH scientific community with network capacity sufficient to meet these research needs via the high-speed FDDI backbone. The FDDI backbone has proven extremely reliable and meets the NIHnet community's current capacity requirements.

NIHnet utilizes conventional telecommunications technologies and infrastructure in order to meet the flexibility and off-campus networking requirements more typically associated with the NIH administrative community. Over 100 LANs are connected to NIHnet using 1.5 megabit per second (T1) telecommunications lines. T1 communications can usually be transmitted over existing telephone lines, which eliminates the need for additional wiring and makes possible timely and cost-effective

installation of new NIHnet connections. In addition, T1 lines are well suited for off-campus connections (e.g., the Westwood Building, NIEHS in Research Triangle Park, NC, the various NIH locations on Executive Boulevard, Parklawn Boulevard and Twinbrook Parkway, Frederick Cancer Research and Development Center, and NIDA and NIA-GRC in Baltimore.

The NSB is working with the NIH Telecommunications Office to use already installed single-mode fiber connections to the buildings on campus currently served by T1 lines. The pilot project to connect Building 6 to the FDDI backbone using single-mode fiber was successfully completed in FY93 and is being used as the model for connecting the remaining on-campus buildings in an orderly manner onto NIHnet's fiber backbone.

One of the most heavily utilized T1 links connects NIHnet to SURAnet's office in College Park, Maryland, the regional component of the Internet. All NIHnet traffic to the Internet, including file transfers and interactive sessions on distant machines, goes through this T1 line to SURAnet. In recognition of the high-capacity data transfer currently being done on this line and in anticipation of even heavier traffic in the near future, the NSB is investigating the use of newly available commercial connectivity solutions for a higher speed link to SURAnet. In FY94, a link to SURAnet with 10 megabits per second capacity will replace the current T1 (1.5-megabits-per-second capacity) line. Based on the success and the economics of this high-capacity solution, other T1 lines to off-campus locations (such as to the Executive Plaza complex) are candidates for replacement with higher speed connections.

NIHnet Monitoring and Support

The Network Systems Branch extensively monitors NIHnet to ensure continuous and reliable service for users on connected LANs. NSB staffers have network monitoring stations, software LAN analyzers, and hardware diagnostic tools to assist in this process. LAN coordinators are contacted in the event of network problems to assist in the diagnosis and solution of the problems.

The network monitoring stations used by NSB display a map of NIHnet, showing correctly functioning connections in green. If a connection to a LAN stops communicating to NIHnet, the network monitor sounds an alarm and changes the depiction of the connection on the network map to red. In this way, NSB staffers are alerted immediately to problems, and can contact the LAN coordinator and start the problem diagnosis and resolution process very quickly. An additional benefit of the NSB's proactive stance is that the NSB staffer's call often gives the LAN coordinator a "heads up" warning that a local LAN problem may be occurring. The NSB is working to further enhance and expand these network monitoring capabilities. NSB provides mechanisms for LAN coordinators to initiate NIHnet questions or problem reports. Perhaps the most popular is the NIHnet Customer Support "hot line," a telephone consulting service available from 8:30 a.m. to 4:30 p.m. each workday and staffed by senior networking professionals. The NIHnet hot line has been in existence since March 1993. On a typical day, hot line staffers will receive between 10 and 15 calls on topics such as LAN wiring recommendations, GRATEFUL MED® installation, Internet Protocol address registration in the Domain Name Server, electronic mail addressing, how to connect to a UNIX® server, and how to determine if a PC is correctly communicating over the network.

The NSB also encourages problem reports or questions via electronic mail, to the NIHNET@LIST.NIH.GOV address. In this way, a LAN coordinator can send complete problem documentation directly to the NSB for assignment to the most appropriate networking expert.

The NSB monitors and promotes the Technical LAN Coordinator interest group, which is accessible via the TLC-L@LIST.NIH.GOV e-mail address. LAN Coordinators can communicate among themselves and NSB can forward information to the TLC community via this list.

Regular meetings with the LAN coordinators, called "TLC meetings," are another way that the NSB communicates with the NIHnet user community. In FY93, the TLC meeting was combined with the Campus User Research Exchange (CURE) meeting, which has a very similar network-based audience. The TLC meeting provides a forum for the NSB and other DCRT groups to make direct contact with networking leaders from throughout NIH for an open and public exchange of ideas, comments, questions, and concerns.

In FY93 and FY94, the NSB is consolidating the inventory of NIHnet components, contacts, and events into a comprehensive database. A single server will host a trouble ticket system for problem and incident tracking; a database of routers, lines and LAN connections; a sophisticated network management and monitoring software package; and a database of LAN coordinators and other network contacts. This consolidation will allow more-effective tracking of incoming calls to the NIHnet hot line and will allow better statistical analysis of network performance, capacity, and reliability.

Preparation for the Unexpected

Providing a production, reliable, sustainable and supportable networking infrastructure for NIH requires preparation for constantly evolving network demand. The NSB regularly deploys Uninterruptable Power Supplies (UPSs) for hub and outboard routers; in FY93 the UPSs for the T1 hub routers were upgraded for additional power protection. In this way, transient power spikes or drops will not knock out or damage NIHnet routing equipment.

Other steps are taken to prevent, minimize, and diagnose network outages: the FDDI backbone is backed up by an ethernet in the event of a ring failure; backup T1 interfaces are configured on the hub routers for quick recovery from hardware failures; routers are capable of being diagnosed via telephone connections in the event of network problems; backbones and redundant connections are in place in Building 31, the Westwood building, and the Executive Plaza North/South complex to prevent outage due to a single T1 line failure. All of these precautionary designs were utilized during FY93 when handling NIHnet events.

FY94 will see the NSB continuing to lead the effort to develop a risk-resistant networking infrastructure for NIHnet. NSB is evaluating the deployment of new routers in the Clinical Center that provide "hot-swappable" components, which will further reduce any downtime due to hardware problems.

Network Guidance and "Whole Building" Wiring

From the LAN perspective, management is easier, more effective and economical, and reliability is higher when the LAN wiring is well designed and properly installed. From the wide-area network perspective, NIHnet is more robust and easier to support if connected LANs are well managed. In light of this, the NSB provides extensive consulting help for NIH groups taking their first step into networking. NSB staffers, in a team approach with the Distributed Systems Branch (DSB), provide "organizational consults" to help define network technologies, strategies, and applications to groups who are preparing to get networked. As a result, when an NIH scientist comes to DCRT with the question, "We want to get our lab networked and participate in NIHnet – how do we get started?", there are DCRT resources in place to assist with suggestions and to provide hands-on assistance.

Organizational consults and advice on network wiring strategies are provided for large as well as small groups. For example, in FY93 the NSB consulted on networking issues with an NIAAA contingent of over 150 staffers moving to Executive Boulevard. The largest groups to get NSB wiring advice have been whole buildings – in FY93 the NSB participated extensively in the wiring plans for Buildings 49, 6, 37, and the upcoming Natcher Building. By providing networking assistance at the outset, NSB encourages standardization in the

network wiring plant at NIH and, as a result, helps ICDs avoid expensive retrofitting and rewiring.

The DCRT Network

The largest single LAN on the NIH Campus is DCRT's Ethernet. Spanning buildings 12, 12A, and 12B, the DCRT Ethernet serves nearly a thousand nodes, including PCs, Macintoshes®, UNIX® workstations, printers, file servers, and mail servers. NSB has the responsibility for maintaining this network – which occasionally proves to be difficult due to the aging cabling plant in place in the Building 12 complex. Reconfiguration efforts are under way to isolate production NIHnet-wide services (such as the Convex and the primary Domain Name Server) from other users of the Ethernet. NSB plans to rewire Building 12A with industry-standard 10baseT cabling as the long-term solution for meeting the networking requirements of DCRT.

NSB Provides NIHnet-Wide Network Services

In addition to planning, deploying, and supporting the NIHnet infrastructure, NSB also develops information resources and network-based applications. These network services add value to NIHnet for the NIH community.

Electronic mail gateways provide one of the most essential network services, because the gateways allow users to correspond across the network via electronic mail with colleagues and collaborators. NSB supports two widely used electronic mail gateways – the 3+Mail gateway and the Microsoft® Mail gateway.

The 3+Mail gateway is used by 70 LANs on NIHnet, passing electronic mail between LANs using the 3Com 3+ operating system and other electronic mail systems. This configuration includes a 3+Mail hub for passing mail from one 3+ LAN to another 3+ LAN. Despite 3Com Corporation's departure from the LAN operating system and electronic mail market, 3+Mail still has an extensive installed base at NIH. On an typical day in FY93, the 3+Mail gateway passed over 1,500 messages among 3+ LANs and to other mail systems. While it is clear that 3+Mail usage at NIH is destined to dwindle, NSB will continue to support the 3+Mail gateway to meet NIH-wide LAN requirements.

DCRT recommends Microsoft® Mail as the replacement for 3+Mail and supplies MS Mail server software to NIH LANs. NSB and DSB have collaborated to install and test the Microsoft® Mail gateway, which converts mail from an individual LAN into the Internet mail standard format known as the Simple Mail Transport Protocol (SMTP) for transmission to other LANs, NIH UNIX® or mainframe users, or remote Internet sites. The initial LAN connections to the Microsoft® Mail gateway were put into place in FY93; the MS Mail gateway, which is currently considered in "pilot production" phase, currently handles mail from 20 LANs at NIH. A number of enhancements are planned for the MS Mail gateway during FY94, including directory synchronization between servers, user address exchange with the NIH e-mail directory, "bullet-proof" backup and recovery systems, and additional operational monitoring.

It is anticipated that the MS Mail gateway will handle mail for the majority of 3+Mail LANs as they convert to Microsoft® Mail. The NSB is working closely with other DCRT groups, NIH-wide LAN representatives, and the vendor to ensure that the MS Mail gateway is an effective long-term electronic mail distribution mechanism for cross-platform and trans-organizational communication.

Several other e-mail-related efforts are under way. In the next year, we will begin a pilot study of routing (rather than tunneling) the Novell® IPX protocol. Assuming that this is successful, we will begin to offer a progressively enlarging suite of services and support for this system. This is necessitated by the significant installed base for this operating system in a few ICDs, its growing market nationwide, and a number of recent technical improvements. NIH will be living with a complex, heterogeneous set of network protocols.

NSB is working on providing an NIHnet-wide fax gateway to replace the 3+Mail-based fax gateway currently on PUBnet. A fax gateway accepts e-mail sent from an NIHnet-connected workstation and converts it to a fax to be sent over conventional telephone lines to the recipient's fax machine. This service makes it possible, for example, to include fax destinations in the CC list for a piece of electronic mail, sending the e-mail via fax machine to people who are not connected to the network.

Dial-up access to the network will be an area of interest for NSB during FY94. In the past year, the technologies for effective dial-up network access have matured. PPP (Point-to-Point Protocol for TCP/IP), ARAP (Appletalk® Remote Access Protocol for Macintoshes®), and Xremote are dial-up network protocols being used in server and client software; modem speeds have increased and costs have decreased; and communications servers have become more common and cost effective. The goal of the NSB is to work with other DCRT groups to provide network access to NIHnet users who happen to be away from their regular NIHnet-connected workstation. Dial-up access is not intended to replace workstation connections to NIHnet, but rather to augment those connections and to add flexibility to the NIHnet access mechanisms available to the user community.

NSB collaborates with other components of DCRT to help provide other network services: with CBMS on *Gopher™*; with CFB on the development of a comprehensive NIHnet e-mail directory; with DSB and CFB on the MS Mail-to-e-mail directory information swapping; and with CSB on a network-based problem tracking system for use by the division and the NIH community.

Technology Tracking for Today and Tomorrow

Computer networking is a volatile industry, with new networking companies and technologies and even philosophies appearing almost daily. The race between new products with higher capacities and new applications with higher capacity requirements is a perpetual dead heat. In light of this, NSB actively tracks and influences networking trends, technologies, and standards. The goal of this effort is to maintain and augment the capacity and sustainability of NIHnet. This must be done within the framework of current technologies while preparing for deployment of upcoming technologies.

A good example of this is NSB's recommendation that NIHnet LANs cable for ethernet in a star and hub "10baseT" configuration using Type 5 wire, which provides 10-megabits-per-second capacity. This conforms to the current industry standard – with benefits including very low cost for ethernet cards and better diagnostic maintenance. In addition, this recommended configuration positions NIHnet LANs for the upcoming 100 megabits per second ethernet technologies that are only now being debated in the networking standards groups. This will allow NIH LANs to leverage their current investment when implementing the next-generation technology. NSB is closely tracking the 100-mbs ethernet technology as a mechanism for handling high-bandwidth scientific applications like medical imaging, full-motion video, and 3D computer modeling. NSB anticipates undertaking a pilot test of 100-mbs ethernet during FY94.

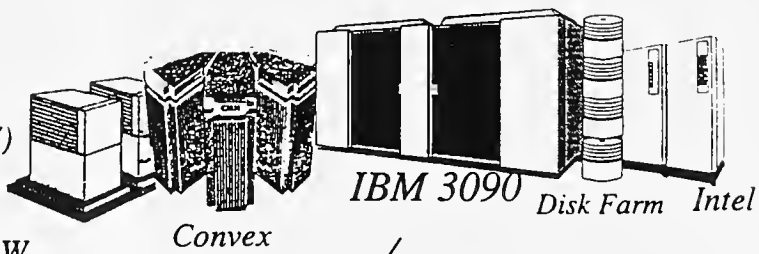
The NSB employs state-of-the-art FDDI technology on NIHnet's intra-campus fiber backbone. The FDDI standard has been ratified by international standards bodies and is implemented by a number of networking equipment vendors. As a result, market forces decrease the cost of FDDI implementation, to the benefit of NIHnet users. As an additional benefit, the fiber infrastructure in place for FDDI will also be compatible with the next generation of backbone technology – Asynchronous Transfer Mode (ATM). The NSB is working with DSB to prototype an ATM network application during FY94 in anticipation of ratification of the ATM standards and commercial availability of standards-based ATM hardware and software. By deploying a standards-based FDDI backbone now and tracking the ATM technology, the

NSB is positioning the NIHnet backbone for tomorrow's technologies and tomorrow's challenging applications.

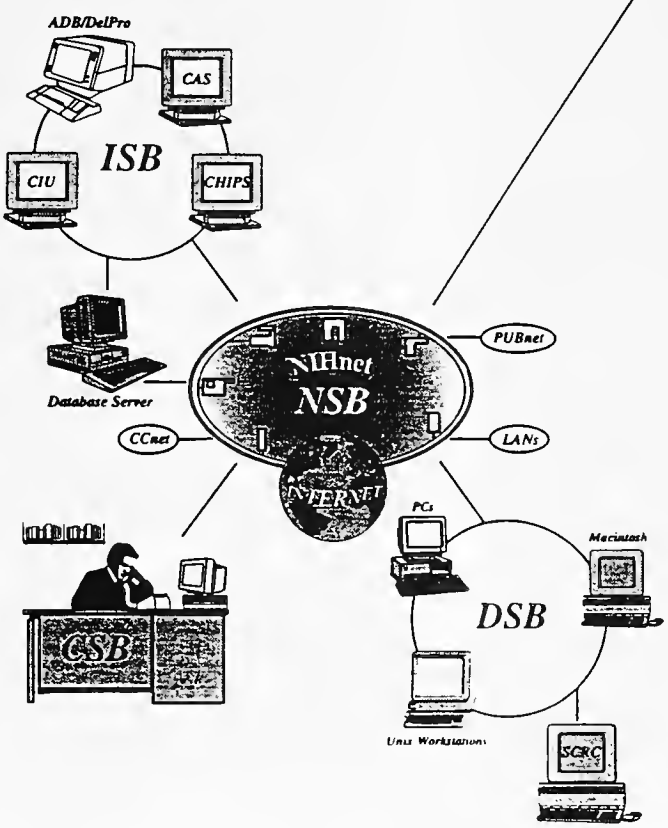
In FY94 the NSB will also be tracking network-based applications in order to forecast NIHnet connection and capacity requirements in upcoming years. Just as the introduction of a network-capable version of NLM's GRATEFUL MED® software was the driving force behind many new network connection requests in FY93, other applications will provide impetus for the NIH scientific and administrative communities to get networked. For example, as client/server database applications are implemented, such as the Division of Research Grants' Information for Management, Planning, Analysis, and Coordination (IMPAC) system replacement to track NIH scientific grants, vast new audiences are likely to want to get connected to NIHnet. NSB fosters substantial personal contacts with the various communities throughout NIH in order to keep a "finger on the pulse" of the network requirements for upcoming applications.

CFB
Computing Facilities Branch

Mainframe - IBM 3090
Disk Farm
Supercomputers
Convex (vector)
Intel (massively parallel)
Central Support of Unix Workstations
Advanced Laboratory Workstations (ALW)
ALW file servers
Workstation Clusters



ALW
File Servers



Computing Facilities Branch

Perry S. Plexico, Acting Chief

The Computing Facilities Branch (CFB) plans, implements, maintains, operates, and supports centrally owned or administered computing resources for NIH enterprise by both scientific and administrative programs. The branch also strives to achieve interoperability among the resources it provides and between them and other computing facilities owned by other organizations in the NIH community. CFB grew out of the Computer Center Branch, which for the past 26 years has provided computing and networking services to NIH research investigators and administrators who conduct and manage modern biomedical research.

The NIH Computer Center and its associated telecommunications facilities are among the resources for which CFB is now responsible. The Computer Center is made up of two interconnected multicomputer facilities designed around large-scale IBM® mainframe and Convex superminicomputers. CFB also has responsibility for DCRT's Advanced Laboratory Workstation (ALW) project, which represents an open-systems approach to distributed computing systems. A centrally administered distributed file system supports UNIX® workstations connected to NIHnet. In the future, CFB can build on these technologies to provide full interoperability among the branch's computing resources, and between those resources and user-owned personal computers and workstations.

CFB provides interactive timesharing, database management, graphics, batch, and high performance scientific computation services on its IBM® mainframes to approximately 17,000 authorized users at NIH and in other agencies throughout the federal government. These services are provided to users on a fee-for-service, full cost-recovery basis. Scientific computing services on the Convex system are funded separately by the NIH Management Fund; Convex usage is restricted to NIH staff. ALW services currently are funded through the NIH Management Fund as well.

Thousands of NIH users on local area networks (LANs) can access all major facilities – IBM®, Convex, and ALW – over NIHnet, the campus-wide area network, in addition to using traditional telephone connectivity. Mail gateways allow central facility and LAN users to interchange electronic mail among themselves and with others via the Internet and BITNET international networks. In addition, services of the Computer Center are accessible to other federal users worldwide through the Internet. The Federal Telecommunications Service (FTS), international 800 service, and commercial switched telephone lines also provide access.

Following a reorganization that took place during FY93, CFB provides its services through the Office of the Chief and six sections:

The *Office of the Chief* sets branch policy, guides strategic planning and direction setting, and manages CFB activities by coordinating the work of the sections to encourage and ensure appropriate cooperation and integration of efforts. The Office of the Chief also exercises principal responsibility for capacity management, disaster recovery, financial management, procurement, and property management functions.

The *Database Systems Section* (DBSS) evaluates, implements, and supports central database systems and tools. Its responsibilities include IMS®, DB2®, and various database servers and gateways that it will introduce to help users implement client/server database environments.

The *Distributed Systems Section* (DSS) investigates and evaluates distributed computing technologies, and applies these to develop and implement networked, interoperable, open-architecture, distributed computing environments. DSS has principal responsibility for the Advanced Laboratory Workstation project.

The *Enterprise Systems Development Section* (ESDS) plans, manages, supports and coordinates hardware and software systems integration and development work related to the IBM® facility and for other platforms that CFB introduces in the future for corporate use on an NIH-wide basis.

The *Enterprise Technologies Section* (ETS) identifies, evaluates, documents, and supports products (other than database) to address the software capabilities needed on the current IBM® 370 platform, and for NIH-wide corporate use with future open-systems architectures. ETS will also serve as the principal link between CFB and the new Customer Services Branch (CSB), providing second-level user support for enterprise systems. Currently, ETS also provides primary user support for the IBM® 370 platform pending a future transition of this function to CSB.

The *High Performance Scientific Computing Section* (HPSCS) plans, manages, and supports centrally located high performance computers specifically designated for scientific use at NIH, and works toward incorporating them into an NIH distributed computing environment. This section currently supports the Convex system.

The *Systems Operations Management Section* (SOMS) manages the maintenance and operation of central computing facilities, the physical plant supporting them, and their physical security. It also provides operations and maintenance support for the NIHnet campus network and offers plotting and output distribution services. SOMS has principal responsibility for introducing automated operations into the Computer Center, including evaluating, implementing, and managing software tools and robotics.

Highlights of the Year

Under the major restructuring of the Division of Computer Research and Technology which took place this year, the Computer Center Branch became the Computing Facilities Branch. Network support previously provided by the Computer Center Branch was transferred to the new Network Systems Branch in March 1993. The Computing Facilities Branch continues to operate the NIH Computer Center and has taken on increased responsibilities for developing, operating, and supporting all centrally owned, shared-use computing resources.

The Advanced Laboratory Workstation (ALW) System, an open, distributed system giving biomedical researchers "plug and play" capability for several types of UNIX® workstations via NIHnet, received the 1992 Best of Open Systems Solutions (BOSS) award in the Innovation in Hardware, Software, and Networking Approaches category. This prestigious award is conferred annually by the Federal Computer Conference (FCC) and the Government Open Systems Solution Council to recognize federal, state, and local government agencies that have best applied open-systems technology. ALWs were showcased at FCC OpenNet '92 national convention, and NIH received an impressive green marble obelisk trophy, now on display.

Three new searchable databases, Current Contents® (SCISEARCH®), REFERENCE UPDATE®, and the Current Index to Statistics, were added to *Gopher™*. A distributed service hosted on the NIH Convex system, and available from PCs and UNIX® workstations at NIH for browsing, searching, and retrieving information on many computers around the globe, *Gopher™* is distinguished by its ability to access remote sites transparently by "tunneling" transparently through the Internet. Current Contents® provides *Gopher™* users with a complete bibliographic record for each item in the table of contents of current issues of over 4,500 leading journals in the sciences. REFERENCE UPDATE® provides a similar service with access to abstracts. The Current Index to Statistics permits searching the online version of 17 volumes of the Current Index to Statistics (1975-1991).

Clustered computing provides high-performance computing by use of networked workstations. A group of networked workstations, potentially located in disparate sites, is managed by a common set of software that distributes incoming numerically intensive compute jobs among those machines. Clusters may consist of dedicated compute servers, personal workstations, or combinations of both. Including high performance personal workstations in the clusters provides the

ability to take advantage of unused cycles. A cluster can support both serial batch and parallel jobs.

This year we selected a queue management software package based on user interface effectiveness, simplicity of administration, availability of support, and the ability for workstation owners to set criteria such as time of day or machine load that must be met before a job can run on their machines. The ability to let workstation owners specify the conditions under which jobs can run on their machines was deemed essential for a cluster that included personal workstations belonging to others. We set up a cluster comprising six IBM® RS/6000 machines and two SUN® SPARC® workstations.

A number of test users running numerically intensive computing kept the six RS/6000s between 20-30% busy for a period of 3 months (April-June 1993). Efforts to recruit more users with numerically intensive jobs are under way. We plan to expand the cluster with additional SUN® workstations as well as other architectures (Silicon Graphics®, Inc.; Hewlett-Packard). Other plans call for integrating the cluster with the Andrew File System®, allowing user file access from all cluster machines. We also expect to consider packaging software for distribution to end-users to help them set up clusters in their own laboratories.

Evaluation of client/server database technologies that can provide connectivity from desktop workstations (PCs and Macintoshes®) to DB2® (the currently supported mainframe relational database server) continued to be an important focus for both the DCRT in general and the Computing Facilities Branch in particular. Throughout the year, a Database Technology Group (DBTG), sponsored by DCRT, met monthly to let users and DCRT staff exchange information on client/server database technologies and related technical topics. One purpose of the DBTG was to identify products that might be used to establish an effective client/server environment at NIH. Meetings focused on defining requirements for both clients and servers, identifying support issues in a distributed environment, and developing a checklist to be used for selecting, evaluating, and purchasing client/server products.

We conducted evaluation and testing of a number of LAN-based relational database servers and gateways during the year, ultimately selecting two client/server products for implementation and user evaluation. The products are SQL*Connect to DB2® from Oracle Corporation, and Net-Gateway™ to DB2® from SYBASE Incorporated. Oracle to DB2® mainframe software has been procured and installed. The Oracle gateway to DB2® is currently receiving level 2 support, and an extended 1-year user evaluation is in progress. A 90-day evaluation of the SYBASE gateway to DB2® is nearly completed; an evaluation report will be released soon.

Planning efforts directed at establishing a disaster recovery plan for the NIH Computer Center continued this year. To encourage contingency planning, a seminar on disaster preparedness was presented for managers and technical leaders responsible for operating computing systems at NIH. The seminar introduced the concepts of contingency planning, and reviewed the specific steps involved in developing a disaster recovery plan. Steps were taken to provide for an alternative "hot site" at which mission-critical applications could be processed should the Computer Center ever become disabled.

Hardware Upgrades Improve Convex Performance

To accommodate the steady increase in system utilization, the NIH Convex system was upgraded in March 1993 to a three-processor C3 series system (C3830) which runs version 10.2 of the Convex UNIX®-based operating system. The Convex has more than 1,700 active users, with up to 120 concurrent users during the day. Before the upgrade, load average, which reflects the number of CPU bound jobs, was peaking at 16 to 20 during the day and remained close to 10 overnight, with many jobs using hundreds or thousands hours of CPU time. The upgraded Convex C3830, which approximately doubled the computational power of the system, has three tightly coupled central processors with an aggregate capacity of 180 MIPS (million instructions

per second) and 360 MFLOPS (million floating point operations per second) in 64-bit mode and 720 MFLOPS in 32-bit mode. The new processors are air-cooled and use gallium arsenide technology, a combination which allows them to run faster and more reliably than systems using older technologies. In addition, an advanced technology called Integrated Distributed Power System allows each board to maintain its own power supply, thereby allowing CPUs to be replaced without shutting down the system. Each of the processors offers hardware vector capability, and the three processors together provide parallel computing via Convex's hardware-based Automatic Self-Allocating Processors technology, which allows a single processor to request additional processors to execute portions of code that can run in parallel. This hardware upgrade, combined with the enhancements and fixes in the new operating system, have increased throughput and responsiveness and improved system reliability and maintainability.

Higher Speeds for Interactive Services

The branch is taking a giant step forward with the introduction of higher communication speeds for users of the Computer Center's interactive services. Network connection of the central facility previously addressed some of the demand for higher speed communications, but did not meet the needs of those without network connectivity. Relief for these users is coming with the introduction of new communication controllers, COMTEN® model 5665s from the NCR Corporation. These controllers provide speeds up to 19.2 kbps (19.2 thousand bits per second) with the possibility of up to 38.4 kbps in the future.

The availability of higher speeds results from using the new communication controllers in combination with newer, state-of-the-art modems. In addition to higher speeds, the new modems also offer error correction protocols, thus reducing transmission errors. Error correction protocols are particularly important when a call originates in an area with poor-quality telephone lines. Faster telecommunication speeds open up capabilities and functions, such as large file transfers, previously not considered viable because of the time delays involved in transmitting large quantities of data.

New DB2® and COBOL Releases Increase Efficiency

A new release of IBM's DB2® was installed in January 1993. Enhancements include more efficient handling of highly unclustered data by the REORG utility, increases in the maximum number of columns in tables and indexes, an increase in decimal number precision, and the ability to create multiple image copies of the same table simultaneously. The new release also offers several new features to assist database administrators and application developers. These features include a facility to view DB2® catalog information and an application programming feature which eliminates additional coding to reposition the cursor.

Three new software facilities also were made available to DB2® users. They are:

- a batch-cataloged procedure to extract data from DB2® and format it for easy downloading and importing into PC and Macintosh® software products
- a QMF procedure that submits QMF commands for execution in the batch
- a QMF procedure that reports DB2® tablespace and indexspace extent information.

In addition, the procedure for printing QMF output was enhanced to permit the printing of a title on header and trailer pages of a job listing and to direct printed output to a queue where it can be accessed through WYLBUR or ISPF®.

A new COBOL compiler was made available for the IBM® System/370 in June 1993, just 6 months after IBM® announced its intention to discontinue support of the OS/VS COBOL® compiler in 1994. Because the COBOL compiler is used more than 14,000 times a month and many of NIH's most important systems are written in the COBOL language, transition to a replacement product was given high priority. The new compiler, COBOL/370,

was chosen after an intensive effort to determine which of the available COBOL compilers would best suit the needs of our users.

Two facilities related to the COBOL/370 conversion were also made available. They are a COBOL Bulletin Board in ENTER BBS and a COBOL-L list on the NIH LISTSERV facility. These provide ways for users to share ideas and problems encountered during the COBOL transition.

Capacity Management and Time Allocation Improve System Utilization

Capacity management efforts begun during FY92 were continued and extended during the past year. The results of these efforts were more cost-effective operation while still ensuring sufficient computing capacity to handle user workload efficiently. Capacity management involves forecasting future data processing needs and workload growth, identifying new user requirements, and assessing future capacity in time to provide expansions, upgrades, or modifications when they are needed. A Capacity Management Staff (CMS), which includes a capacity planner, was appointed to analyze projected workloads and available capacity and project future resource requirements. One result of capacity management efforts was the removal of one of the four 3090 CPUs in our MVS production complex, which was accomplished with no degradation in service levels. This was accomplished using PR/SM® to permit the merger of the test/backup machine with a production machine.

In the coming year we plan to institute better management of the NIH Convex system by introducing a technique called time allocation. A Time Allocation Group, comprised of NIH and extramural scientists well versed in the application of computers to biomedical research, will review formal applications for large blocks of CPU time. Evaluation will consider scientific merit, appropriateness of the project to the Convex supercomputer, and the ability of the researcher to make effective use of the system as indicated by preparatory work and/or preliminary results. Most current Convex users will be unaffected by this change because their requirements for CPU time are relatively modest.

ALW Expands Offerings and Prepares for Cost Recovery

New applications introduced for ALWs this year include MLAB (mathematical modeling), SAS® and SUDAAN (statistical analysis), *Gopher™* (multimedia information retrieval), and the UNIX® version of the popular WordPerfect® editor. We began supporting HP® 9000 Model 700 workstations and also Silicon Graphics® workstations (via the NFS/AFS Translator). We upgraded DEC® workstations to Ultrix™ 4.3 and SUN® workstations to SUNOS 4.1.3. All systems now use a new facility called *depot* to manage testing, configuration, and distribution of applications software, thereby greatly improving software quality control and management.

In FY94, we plan to enhance support for Silicon Graphics® workstations by offering direct access to the Andrew File System® (AFS) for Silicon Graphics® machines running IRIX 5.1, to upgrade at least two of our older SPARCserver® 490 AFS fileservers to SPARCserver® 1000s, to upgrade our base of SUN® workstations to Solaris® 2.x, and to develop ALW support for the new DEC® "Alpha" workstation running OSF/1. We also plan to introduce an Environment Management Tool (EMT) with a graphical user interface to enable application maintainers and developers to use *depot* to manage their own software collections.

Over the past year, the ALW system has shifted from being a research and development activity to becoming a service offering. Accordingly, we have made preparations for it to recover one-third of its operating costs in FY94, two-thirds in FY95, and to achieve full cost recovery in FY96. This will be accomplished by charging an ALW installation fee, a monthly workstation subscription fee, a daily fee for disk storage, and a monthly fee for managing privately owned AFS file servers. Current ALW users

have been mailed forms for obtaining DCRT accounts and registering ALW users, client machines, and storage groups, and we have designed and tested billing software to generate records for processing by DCRT's Project Accounting System. Finally, we developed requirements for a contract to obtain technical support services, which we plan to award late in FY94.

Data Management Made More Efficient

We completed the conversion to an all-cataloged environment during FY93, implementing changes that removed the ability to designate specific volume serial numbers when creating or accessing data sets on public disk volumes. This conversion has improved the efficiency of Direct Access Storage Device space usage, reduced the cost of storage, and simplified Job Control Language. Now that the transition to the all-catalogued environment is achieved, we plan to begin concentrating on new ways to improve data management, including the implementation of IBM's System Managed Storage facility.

Cost Savings Passed on to Users

For the 25th consecutive year, we were able to offer significant rate reductions for some services and offer rebates to all users. This year, we began with a 21% discount on all user invoices for the month of October 1992, followed by rate reductions which became effective on November 1, 1992, and provided users with a cumulative cost saving in excess of $850,000 per month. These reductions were followed later in the year with a major rebate and additional billing discounts. The total savings during the year amounted to 55% relative to what it would have been with the rate structure of September 1992.

Data storage, which represents the largest single cost category for most computing activities, received the largest reduction as the rate for public FILE storage was reduced 70%. Other changes reduced the cost of MSS (Management System Storage) storage and dedicated volumes as well. At the same time, the batch processing charging algorithm was simplified to reflect more accurately the costs of the various resources consumed. In March 1993, we distributed a rebate to all users in two parts. First, a $10.4 million refund was distributed to all user accounts in March, prorated according to usage during the first half of the fiscal year. This was followed by a 28% discount applied to each remaining monthly statement in the fiscal year.

Documentation Services

The CFB Technical Information Office (TIO) provides high-quality documentation services to users of the NIH IBM® mainframe and Convex super-minicomputer. The TIO manages orders and inventories to keep hundreds of different documents on hand for users. Publications may be picked up in person, but most users submit their orders electronically and have the publications sent to their offices or placed in their output boxes for messenger pickup. Annually updated subscription profiles ensure that users automatically receive the latest versions of their regular publications. Some publications come from vendors; others are written in-house to describe the use of facilities unique to the NIH computing environment and to keep users up-to-date on the latest changes in the systems. This year, five editions of INTERFACE, CFB's series of technical notes, were distributed: four regular editions and one special issue. Three major updates of the two-volume "Computer Center Users Guide" were issued.

Training Opportunities

The DCRT Computer Training Program continued to provide a diversity of educational opportunities for computer users. Scientific seminars were held on a wide array of topics, including biophysics, biochemistry, molecular biology, computer science, biostatistics, biomathematics and bioengineering. Computing seminars ranged from basic introductions to WYLBUR and DB2® to

CFB SERVICES FOR USERS – FY93*

CATEGORY	TOTALS
TRAINING	
Formal Training Classes Offered (Brochure)	229
Number of Students Accepted into Classes	4,250
Number of Self-Study Courses Offered	30
DOCUMENTATION DISTRIBUTION	
Number of Pieces Distributed	169,792
Number of People Receiving Documentation	5,604
DOCUMENTATION PREPARATION	
Number of New/Updated Documentation Pages Prepared	3,260
Number of New/Updated Documents Published	46
CUSTOMER CONTACTS	
Number of User Contacts (PAL Counter & other calls)	32,158
Number of Consulting Appointments**	126
Number of PTRs Handled	2,490
NUnet SERVICE AND MAINTENANCE†	
Number of LANs Supported	174
HARDWARE AND SOFTWARE SERVICE	
New Software Installed	10
Number of Fixes Applied	15,432
Old Software Upgrade (Vendor Supplied)	22
SECURITY	
Number of Formal Security Investigations	10
Number of Keyword/Password Assistance Cases	546
NUMBER OF USERS	
IBM	18,841
Convex	3,016
Advanced Laboratory Workstations	267

*This includes 2.5 months that the Training Program was part of the Customer Services Branch.
**Consulting Appointments were discontinued as a separate service in June 1993.
†Up to March 1993, after which the Network Systems Branch became responsible for this function.

Selected Courses and Seminars in the DCRT Computer Training Program Coordinated by CFB*

Seminars

Mainframe Services at NIH
Cluster Computing
Molecular Graphics: Creating Pictures and Videos
Analysis of Ligand Binding Data Using the LIGAND Program
Software Engineering and CASE Concepts
Networking for the Scientific Community
Enter BBS

PC

Getting Started with Windows
A Look at DOS 6.0
DOS Batch Files

Macintosh

Macintosh/PC Data Exchange
Macintosh Networking with TCP/IP

IBM

Designing Tables and Managing a DB2 Database
Developing Data Entry Applications With SAS/FSP
Beyond Basic WYLBUR

UNIX

UNIX Commands
Andrew File System
Introduction to the Convex Supercomputer
Gopher
Getting Started with C

*This includes 2.5 months that the Training Program was part of the Customer Services Branch.

technical presentations on parallel processing, presentation graphics, and UNIX® commands.

Classes offered for the first time this year included "Getting Started with Windows™" and "Macintosh®/PC Data Exchange." The increasing demand for dataset protection led to the introduction of a course titled "Data Security Using RACF." Other new courses included "Distributed Database Processing Using Client/Server Technology" and a four-part series called "Managing Information: the Database Paradigm."

Self-study courses, which utilize texts, workbooks, and interactive practice exercises, added depth to the curriculum. Video self-study courses were also available. This year a six-module independent self-study series on "LAN-WAN Internetworking" was introduced.

In July 1993, the Training Program was transferred from CFB to the newly established DCRT Customer Services Branch. This change should ally the Training Program more closely with other user services representing programs from throughout the division, and should serve to simplify and streamline access to training in the future.

A Look Ahead

The reorganization of DCRT will continue in the coming year, with a new Customer Services Branch (CSB) taking over some direct user service responsibilities, such as those currently provided by the PAL Unit.

The Computing Facilities Branch will undertake a number of new strategic directions that

align with DCRT's strategic plan. For example, under the reorganization, CFB has assumed responsibility for DCRT's ALW project, which represents an open-systems approach to distributed computing systems. A centrally administered distributed file system supports UNIX® workstations connected to NIHnet. In the future, CFB can build on these technologies to provide full interoperability among the branch's computing resources, and between those resources and user-owned personal computers and workstations.

CFB plans to offer automated storage management services to users of personal computers, workstations, and local area networks. Backup/restore and archive/retrieve services will be provided via high-speed network connections. This will utilize the storage capability of the mainframe more effectively and prevent user losses when PC or server disk drives are damaged or destroyed. With storage management services, CFB can also facilitate the distribution of software to personal computers and workstations automatically (after appropriate software licensing issues are resolved).

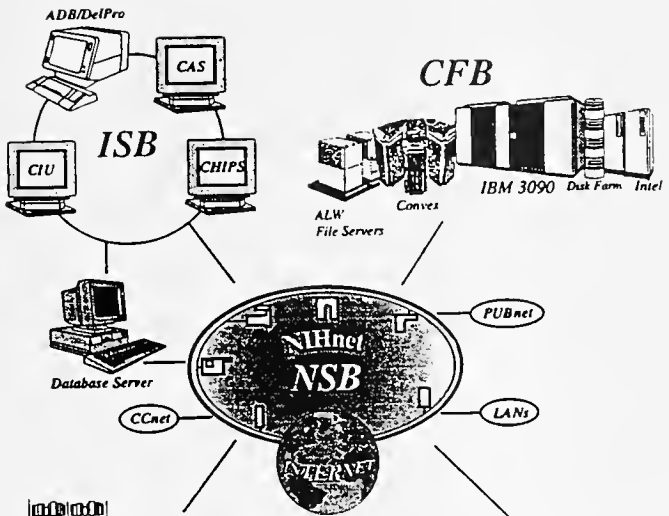
Future expansion of high performance computing offerings will include massively parallel computing, cluster computing, and other technologies as they develop. High performance computer tools are immensely valuable to scientists who apply them to numerically intensive problems in fields such as structural biology, image processing and computational chemistry. As NIH administrative applications of these tools emerge (e.g., for database searching), CFB will provide access to them as well.

Support for large corporate databases that require the reliability, availability, and security of centrally managed facilities will continue, but we expect to enhance their value and usability by making them accessible using emerging client/server technologies. New hardware and software platforms for NIH-wide information management will emerge, which will require collaborative evaluations by several branches within DCRT and the ICDs. The distributed computing and open-systems arenas will contribute an expanded array of systems and database management tools.

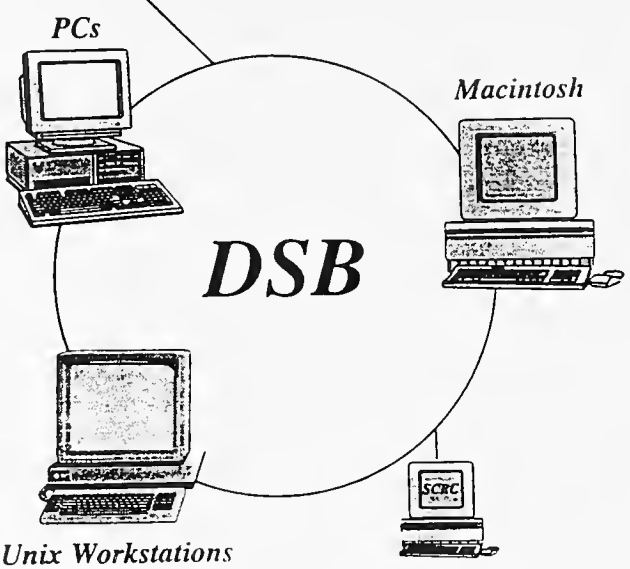
Open systems are those that achieve interoperability, portability, and scalability through the use of generally accepted, widely available standards. By using open-systems technologies, CFB plans to offer users new tools for developing applications or re-engineering old ones. When appropriate, CFB will supplement its proprietary MVS-based system with open-systems interfaces to MVS and with an alternative mainframe system which conforms more closely to open-systems standards.

The dramatic technological shifts facing us as we prepare for the 21st century clearly require organizational change. The restructuring of the organization offers the potential for greater flexibility and growth in the future. The ultimate goal is to offer state-of-the-art computer technologies to address the evolving needs of scientific research and administration at NIH.

DSB
Distributed Systems Branch



PCs & Macintosh
Unix workstations
Application servers



Distributed Systems Branch

David C. Songco, Chief

DCRT services an increasing number of laptop, desktop and benchtop computer users at NIH. This is coupled with a dramatic increase in the number and complexity of the products and applications they use. Although the number of users may level off soon at around 80% of the NIH population, or 12,500 users, the complexity of cooperative processing and distributed computing will continue to challenge DCRT's ability to provide comprehensive support and guidance at the ICD level.

As processing power becomes ever more affordable and user tools become more sophisticated, users are requesting access to many information sources and integration of multiple technologies. Research at NIH depends on an integrated support infrastructure that relies increasingly on computer technology as a unifying instrument. NIH, like much of industry and government, must develop and implement a plan to integrate distributed computing technologies effectively. Design and implementation of new networking and distributed computing architectures is now under way.

The Personal Computing Branch has provided leadership in the establishment and delivery of support for personal computing at NIH since its beginning in 1984, while the Computer Systems Laboratory has supported laboratory and clinical applications, and we are now positioned to provide support for the next chapter in the computing paradigm, networking and distributed computing.

In response to this challenge, the Distributed Systems Branch was formed as part of the DCRT reorganization. The DSB serves as the DCRT focal point for the development, support, guidance, and application of local area network, workgroup, and other forms of distributed computing at NIH. The DSB mission includes advocacy of, advice about, and assistance with the spectrum of computing platforms normally found in the office, laboratory, or clinic.

Most of the staff and function of the Personal Computing Branch were combined with components of the Computer Systems Laboratory as well as staff from two other labs to form the DSB. This integration brings together nearly two dozen highly skilled computer specialists whose mission is the ongoing support of PC, Macintosh®, and LAN users, with about a dozen senior engineers, programmers, mathematicians, systems analysts, and scientists whose mission is the design and development of laboratory and clinical systems in support of biomedical research. In addition, to bring particular emphasis to the special computing needs of NIH scientists, the management of the DCRT Scientific Computing Resource Center was formally assigned to the DSB. The SCRC staff of four supports the computing needs of NIH researchers by providing one-on-one assistance and access to scientific software running on advanced personal computers and UNIX® workstations.

DSB Functions and Organization

The DSB is responsible for assessing the computing requirements of the NIH user community in the area of distributed computing technology and for ensuring that those requirements are addressed in the future goals of DCRT. DSB staff provide guidance and support for NIH organizations in the selection and effective use of emerging computing technology in the laboratory, clinic, and office environments.

The Distributed Systems Branch is organized around two major support areas: scientific computing, and distributed computing and workgroup productivity.

Scientific Computing

The scientific computing functions are organized into four groups:

Clinical Applications Section (CAS)

CAS staff perform evaluation, design, development, and implementation of novel computer systems for clinical signal and image processing problem areas; software for the collection, analysis, and display of physiologic waveforms, high-resolution medical image processing and display techniques; and high-speed network technology in support of medical image transmission, storage, and display. CAS staff also perform requirements analysis and engineering design support for a variety of clinical application areas.

BioInformatics and Molecular Analysis Section (BIMAS)

BIMAS staff provide the NIH community with resources for high-level computational analysis of data in the field of molecular biology. This includes expertise in software development, applying computational analysis techniques to biological data, and providing tools for accessing and displaying large amounts of genomic data from a variety of distributed databases. BIMAS staff are also developing an open "Molecular Biology Workstation."

Scientific Computing Resource Center (SCRC)

The SCRC provides NIH with a shared-use computing facility staffed by computer professionals, where researchers are able to focus on scientific applications. Specialized peripherals for shared use, such as color printers, scanners, and film recorders, are available. The SCRC addresses the needs of the NIH scientific community by providing access to specialized scientific software for image processing, molecular modeling, numerical analysis, sequence analysis, and statistics running on advanced personal computers and UNIX® workstations. SCRC staff also schedule appointments and coordinate additional consultations with DCRT subject-matter consultants. More information on the SCRC can be found in a later section of this report.

Biostatistical Consulting Section (BCS)

The fourth component focusing on scientific computing is the Biostatistical Consulting Section. This group provides a variety of subject-matter consulting services in collaboration with the SCRC.

Distributed Computing and Workgroup Productivity

The PC, Macintosh®, and LAN guidance and support functions previously provided by the Personal Computing Branch are now organized into two sections of the DSB: the System Consulting Section and the System Integration and Development Section. Both sections track and evaluate emerging desktop, workgroup, and local area network computing technologies and document and publish results of evaluations. Staff are involved less, however, in individual user problem resolution and instead are focusing on "organizational consulting": developing system models and recommendations for technical solutions to common NIH computing situations. Staff also collaborate with the Network Systems Branch on the development of campus-wide network services and with the Computer Facilities Branch and the Information Systems Branch in the development and testing of client/server architectures.

The staff of the System Consulting Section serve as the primary DSB interface with the NIH scientific and administrative user community. In this capacity, they assess the technical requirements of the NIH user community in the area of distributed computing, including PCs, Macintoshes®, workstations, LANs, and associated technology and provide guidance in the selection and effective use of emerging computing technology in the laboratory, clinic, and office environments.

System Integration and Development Section staff work closely with the System Consulting Section, but they spend more of their time developing selected NIH-specific solutions by conducting computer science and engineering research, system design, and software development

directed toward the application of distributed computing technologies to NIH programs.

The DSB and the NIH Training Center cosponsor walk-in user resource centers for public NIH access to personal computing resources and information.

Distributed Computing Technology Highlights

Personal computer users at NIH look to the DSB for guidance in the selection of computer hardware and software. Keeping up with the latest developments and predicting future trends is becoming an increasingly challenging task due to the accelerating pace of technological development.

Hardware

Every year personal computers grow more powerful while dropping in price. In FY93, the DSB evaluated and announced support for a number of new systems by Dell™, IBM®, and Apple®, together with a wide range of peripheral options.

On the PC side, the DSB recommended systems based on the price/performance-leading Intel® 80486 processor, and ceased recommending purchase of 80386-based systems. Although Intel® began producing its successor to the 486 chip, the Pentium™, Pentium™-equipped systems remain so expensive and hard to come by that the DSB will wait until FY94 before evaluating and recommending them. Also in FY93, promising alternatives to the Intel® processor line appeared, including Digital Equipment Corp.'s Alpha chip, which is said to deliver performance roughly twice that of the Pentium™. On the Macintosh® side, the forthcoming PowerPC™ processor holds the promise of greater power and PC compatibility. The DSB will closely track alternative processors in FY94.

There was a notable downsizing trend in personal computers in FY93. Leading manufacturers introduced slimmed-down desktop lines and released a plethora of notebook-sized portable systems, many with the power of desktop units. Color PC laptops, a costly rarity in FY92, became commonplace and affordable. The DSB also looked at a number of new "subnotebook" computers, some small enough to fit in the palm of one's hand. These included the Zeos™ Contenda, Gateway™ Handbook, IBM's ThinkPad™ 500, Hewlett-Packard's HP®-100LX, and the Apple® Newton®. The last, termed a "personal data assistant," represents a new class of device altogether, a melding of specialized computer and traditional notepad functions. Pen-based hand-held systems could potentially appeal to many clinical and laboratory workers at NIH, but the continuing lack of standards in the field made it unwise for the DSB to recommend any model.

In response to an executive order requiring Federal agencies to buy energy-efficient computer systems, major manufacturers announced "green" computer lines. Among the first to market such a system was IBM®. However, we held off recommending this system because of some serious design limitations. The DSB announced support for one laser printer meeting government energy guidelines: the LaserJet 4L.

Software

Clearly the most conspicuous trend in the personal computer software industry was the ascendancy of the graphical user interface (GUI), represented mainly by Microsoft® Windows™ and the Apple® Macintosh®, over the traditional command-line interface of DOS. Windows™ appears destined to be the dominant interface for at least the next several years. Microsoft® has reportedly sold more than 25 million copies to date, and Windows™ now sells at the rate of roughly one million copies per month. Most new PC systems sold today come with Windows™ pre-installed. In FY93, for the first time, sales of Windows™ applications exceeded those of DOS applications.

The DSB highlighted Windows™ and Windows™ applications in its publications and presented seminars introducing Windows™. Although we lack precise figures on Windows™

penetration here, we estimate it to be in the 15-25% range and rapidly rising. We believe the main obstacles to its adoption at NIH are underpowered hardware and users with very modest software requirements. Because we feel even those who use their computers mainly for word processing will benefit greatly from the GUI's ease of use features, we intend to redouble our efforts to persuade PC users to buy hardware capable of running Windows™ and to make the transition. Since virtually all future PC software will be developed for Windows™ rather than DOS, DOS users will find themselves using progressively obsolete software technology.

In FY93 Microsoft® released high-end members of the Windows™ family: Windows™ NT and NT Advanced Server. The DSB actively participated in NT beta testing and is currently evaluating the released products. Featuring numerous advanced operating system characteristics, NT is positioned as a server and power user operating system. We expect to recommend it for most new network installations in FY94 and for power users who meet its hefty hardware requirements (386 or 486 processor, 12MB-16MB RAM, 75MB free disk space). To most users, we will continue to recommend Windows™ 3.1 and Windows™ for Workgroups (a peer-to-peer networking version of Windows™ 3.1).

IBM® released OS/2® 2.1 in FY93. We consider it a sound product, in many ways similar to Windows™ NT but with somewhat more modest hardware requirements (386 or 486 processor, 8MB-12MB RAM, 30MB free disk space). Some of its strengths, such as running multiple DOS and 3270 sessions, appeal to certain groups at NIH.

Use of the other mainstream GUI system, the Apple® Macintosh®, appears to have leveled off at about 20%. We consider it a reasonable alternative to Windows™ for groups currently standardized on the Macintosh® platform.

"Groupware," network-based software designed to enhance communication between members of groups, was one of the hottest software genres in FY93, and the DSB actively evaluated groupware products. On the PC side, we looked at a variety of programs for orchestrating members' schedules, including the scheduler that comes standard with Windows™ for Workgroups and Windows™ NT and the one in WordPerfect Office™; sharable personal information managers, such as PackRat™; and group document managers like PageKeeper™ and Folio Views™. We intend to evaluate the leading forms-routing packages in FY94. On the Macintosh® side, we evaluated the Meeting Maker® XP calendar package, an "in/out" board, a group word processor, and a sharable flat-file database called FileMaker®. The last is of special interest since it is available in a Windows™ version as well. Because some NIH sites have a mixture of PCs and Macintoshes®, we consider cross-platform availability a plus in a groupware package. As one potential application for groupware at NIH, we are looking for ways to make NIH resources, such as conference rooms and auditoriums, easier to schedule.

Passing elaborately formatted electronic documents between systems equipped with different word processors and fonts is fraught with problems. In FY93 several promising new programs offered differing solutions to the problem: Adobe® Acrobat™, No Hands Software's Common Ground, and Farallon's Replica. So far, only Acrobat™ is available in both PC and Mac versions. One potential application for such a program is in electronic distribution of the DSB *Product Information Guide*, which we currently provide only in text and selected word processor formats.

Networking

DSB has been collaborating with NSB, CSB, and CFB on various aspects of networking. The number of local area networks at NIH grew during FY93 to over 220. To help meet the increasing demand for assistance and guidance in maintaining these LANs, the DSB established a technical support contract with Microsoft®. This contract gives individual ICDs the option to purchase support that provides direct access to Microsoft® engineers for all Microsoft® networking products. The DSB will be working with the Customer Services Branch in FY94

to expand this contract to include more ICDs and the full range of Microsoft® products.

To prepare for the future of networking at NIH, the DSB established the Advanced Network Operating Systems Project, which is addressing the use, support, and administration of LAN Manager and Windows™ NT Advanced Server. The project team is developing recommendations for all aspects of the implementation of Windows™ NT Advanced Server, including training, administrative procedures, security, domain concepts, and interconnectivity.

There still exists a large number of 3Com 3+Share local area networks at NIH. The DSB has been working with several ICDs in the planning and implementation stages of migrating these LANs to LAN Manager. With the introduction of Windows™ NT Advanced Server at NIH, the DSB expects that most of the remaining 3Com LANs will migrate in the next 2 or 3 years.

DECnet Management

The number of DECnet nodes on the expanding NIH campus-wide network continues to climb. An increasing number of VAX™ sites use DECnet networking protocols to support personal computing via DECnet-based server configurations. Nearly three-fourths of the DECnet nodes now registered are personal computers using work group servers for file and print services. We are continuing to provide the centralized coordination necessary for smooth incorporation of these new DECnet hosts into an integrated network that spans the entire NIHnet campus network.

Electronic Mail

DCRT announced support for Microsoft® Mail as a successor to 3Com's 3+Mail in FY92 after careful consideration of the various electronic mail alternatives. In FY93, the DSB assisted the Network Systems Branch in the development and implementation of a central NIH Microsoft® Mail hub and Standard Mail Transfer Protocol (SMTP) gateway. In addition, the DSB coordinated the distribution of Microsoft® Mail to the ICDs and assisted them in its implementation. There are currently more than 30 LANs running Microsoft® Mail, and we expect the number to grow as networks migrate away from 3Com 3+Share.

The DSB continued to promote the use of TCP/IP (Transport Control Protocol/Internet Protocol) and Appletalk® network-based programs for the Macintosh® and provided consultations on applications in both areas. TCP/IP programs supported include products for remote log-in, FTP file transfer, and various Internet access tools. Appletalk® programs included Appleshare®, System 7™ file sharing, and network-based groupware applications such as Retrospect®, a network backup program that allows workgroups to have local data backed up to a central file server.

In FY93, we also continued our evaluation of X-Windows™ clients and servers for the Macintosh®. Working together with other DCRT groups, we evaluated network-based systems including network analyzers and network server clients. Another major collaboration was with the NIH *Gopher™* team in evaluating and supporting new *Gopher™* clients for the Macintosh® and PC. These efforts will continue in FY94.

Training

Besides offering guidance in the selection of computer products, the DSB, in collaboration with the NIH Training Center (DPM) and the DCRT Training Program (CFB and CSB), provided users with high-quality training in the use of DSB-supported products.

On the DCRT side, there has been a continuing interest in standard DOS courses (Intermediate DOS; DOS Advanced Topics; Batch Files). In addition, there has been greatly increased DSB activity in presenting 2- and 3-hour limited-focus seminars.

Those of particular interest to the scientific community were:

- Comparing Macintosh® Sequence Analysis Programs
- Multiple Sequence Alignment
- Preparing Figures for Publication on the Macintosh®
- Using Computers to Find Possible Regulatory Elements
- Bibliographic Manager Programs for the Macintosh®.

Some of general interest to computer users were:

- Mac and PC Viruses
- A Look at DOS 6.0
- Memory Management on the PC
- PC/Macintosh® Data Exchange
- Macintosh® Networking with TCP/IP
- Windows™ Applications Strategies
- OS/2® Overview.

On the NIH Training Center side, there has been a wealth of activity in DOS-based courses:

- WordPerfect®
- Lotus 1-2-3®
- dBASE III® and IV®
- Paradox®
- Disaster Recovery.

Windows™-based courses included:

- WordPerfect®
- Lotus 1-2-3®
- Excel
- Harvard Graphics™
- PageMaker™
- Microsoft® Project.

Among the Macintosh® courses given were:

- WordPerfect®
- Microsoft® Word
- Excel
- Lotus 1-2-3®
- FileMaker®
- 4th Dimension®
- PageMaker™
- QuarkXpress®
- MacDraw®
- KaleidaGraph™
- PowerPoint®
- HyperCard®.

A networking course (3Com) was also given. Training figures for FY93 showed that 326 personal computer course sessions were presented by the NIH Training Center. Of those, the DSB cosponsored 74 sessions of 11 different courses, attended by 1,200 students. Additionally, 13 sessions of 3 courses (208 students) and 44 seminars were presented through the DCRT Training Program without fee. DSB staff taught 25% of all courses and seminars, and provided direction, course materials, and assistance to enable outside vendors to teach the remainder.

By looking at enrollment figures, the trends seem to evidence a waning interest in some old DOS-based standbys such as WordPerfect®, Lotus 1-2-3®, and especially dBASE; and a great surge in courses on Windows™-based applications and courses with specific appeal to the scientific community.

As in previous years, the DSB Associate Instructor Program played a large role in the success of our training efforts. Under the program, experienced NIH computer users volunteer their time assisting the primary instructors during hands-on training courses. During FY93, 53 persons from 18 ICDs and the Office of the Director participated in 120 course sessions.

In addition, the NIH User Resource Center (URC) continued to serve as a vital adjunct to DSB training services. Sponsored by the DSB in collaboration with the NIH Training Center, the URC is a multipurpose walk-in computer facility equipped with PC and Macintosh® workstations, an extensive collection of applications software, and a variety of peripherals such as laser printers, modems, CD-ROM players, and page scanners. It also has a large selection of self-study courses and many popular personal computing periodicals, books, catalogs, and other publications. During FY93, NIH employees made more than 6,072 visits and 2,410 telephone calls to the URC for such purposes as researching computer topics, evaluating DSB-supported hardware and software, taking self-study courses, and consulting with URC staff. In addition, a new User Resource Center opened at Executive Plaza South to serve the needs of the NIH community in the Executive Boulevard corridor. Remaining popular was the URC Learning Assignment Program, under which NIH employees volunteer 4 hours of their time per week for 3 months in exchange for the

opportunity to enhance their computer skills by working directly with URC staff.

Consulting Services

The DSB maintains a telephone help line during normal working hours. When users call, a dispatcher notes their problem and either puts them through to a specialist immediately or takes their name and number and sends the appropriate specialist an electronic mail message. Consultants generally return calls within a half hour. Alternatively, users can send queries to the dispatcher via e-mail or post questions on the DSB's electronic bulletin board system. During FY93 the DSB responded to more than 5,000 requests for assistance. Enhancements made to the consulting system shortened overall response time, provided more accurate statistics, and allowed us to compile responses for future reference. The DSB significantly increased its use of vendor-supplied help-desk support, thereby freeing staff for more specialized organizational consultations.

The DSB looks to its Lead Users and Macintosh® Support Coordinators to handle routine support problems at the local ICD level. Numbering about 200 and 90 respectively, these persons receive free DSB-sponsored training and priority access to DSB staff in return for the support they provide their own organizations. Lead users meet monthly for DSB presentations on topics of interest. Increasingly in FY93, the DSB relied on vendors to demonstrate their own products at lead user presentations. For example, IBM® and Microsoft® did side-by-side introductions to OS/2® and Windows™ NT, and Lotus Development Corp. showed Notes™. Also in FY93, the DSB took the first step toward bolstering its training program by underwriting most of the cost of a comprehensive vendor-taught PC troubleshooting course for select lead users. The DSB plans to offer more such courses in the future. By increasing the level of computer expertise in the ICDs, we hope to further reduce our support burden.

The DSB also sponsors and supports several user groups across campus as a further means of promoting greater self-sufficiency among users. The Campus Users Research Exchange (CURE) meets monthly for network-related presentations and information exchange. In FY93, it merged with the NIH Technical LAN Coordinator program, a group of NIH LAN administrators. Also active was the Biomedical Research Macintosh® Users Group (BRMUG), which featured monthly presentations on such topics as digital imaging and demonstrations of popular new products like Word, Excel, and MacWrite®. The DSB continued to provide expertise and guidance to the WordPerfect® Working Group. This group is led by members of the National Heart, Lung, and Blood Institute and the Office of the Director and meets monthly for presentations.

Information Dissemination

In addition to relaying timely computer-related information to key NIH persons through Lead Users, Macintosh® Support Coordinators, and user groups, the DSB continued its formal vehicles for NIH-wide distribution of computer-related information: *PCBriefs*, PUBnet, PCBull, and the DSB *Product Information Guide*.

In FY93, five issues of *PCBriefs*, the DSB's technical newsletter for NIH microcomputer users, were published and distributed to 8,000 NIHers. Also in FY93, we published several editions each of the PC and Macintosh® versions of the DSB *Product Information Guide*, which lists and describes all DSB-supported hardware and software products.

PCBull, the DSB's electronic bulletin board service, remained popular in FY93, averaging 700 calls per month. Accessible 24 hours a day from personal workstations via telephone communications, PCBull is an important part of the DSB support program, offering NIH users information, software updates, virus protection, and tips on using DSB-supported products, as well as DSB utilities, publications, and useful public-domain files for the PC and Macintosh®. In addition, users can ask questions online.

Increasingly in FY93, information was distributed to NIH computer users through electronic

mail groups. As the number of LAN users increases, we expect this channel to become even more useful. As always, the User Resource Centers played a large part in the dissemination of DSB information.

The DSB also organized the collection and distribution of electronic forms in FileMaker® Pro, Perform Pro, and WordPerfect® formats. The new forms were distributed on PUBnet, the public network run by the DSB. Thus a number of complex forms used on the NIH campus became electronically available to both Macintosh® and Windows™ users. These electronic forms print out exactly like the forms in use at present, allowing a crisp, easily readable copy as well as maintaining an electronic record of how the forms have been used within a group. The response from the NIH community to the new forms has been very positive, and we expect to expand our collection of NIH-relevant forms in FY94.

In FY93, the DSB held its first annual Macintosh® Workgroup Productivity Show to demonstrate low-cost Macintosh® computers, group computing, and cross-platform connectivity, and the general administrative usefulness of the Macintosh® platform. The DSB was also represented in the Fall at the NIH Research Festival, with posters on such topics as DSB services.

Other Support Services

VAX/VMS™ Minicomputer Services

Currently there are approximately 100 VAX/VMS™ systems at NIH, both on campus and off, that provide computing services to nearly 1,000 users. In recognition of the importance of these systems to the NIH scientific community, DSB continues to enhance our VAX™ and VMS support. These new hardware capacities and enhanced software capabilities will ensure our ability to support our user community as they migrate to expanded capabilities as well.

This activity provides hardware and software support resources for both VAX/VMS™ and AXP/VMS computing operations throughout the NIH. Software tools include language compilers, database systems, consultation services, and connectivity products. The three-member VMS cluster has been enhanced by upgrading the MicroVAX II™ to a VAX™ 4000-100 with 80 Mbytes of memory and an additional 7 Gbytes of disk space and by the addition of a DEC® AXP workstation. Supported software includes FORTRAN, Pascal, C, and C++ compilers, several code development support utilities, network diagnostic utilities, distributed file and queuing service software, relational database development tools, screen and hardcopy graphics support libraries, and a large library of user-contributed software. We are now adding personal computer file and print service support via the Pathworks product suite. Consultative services are still furnished primarily through the contractor-operated "VMS Hot Line" as an adjunct to consultation provided by DSB staff.

The major enhancement to the VMS support effort has been the acquisition of a DEC® 3000-400 AXP RISC-based workstation as the heart of a VAX-to-AXP VMS migration assistance resource. This system, which is equipped with several compilers and code development and migration tools, will enhance our ability to assist our user community as they migrate their applications to newer computing platforms.

DCRT LANs

The DSB administers the DCRT local area networks in the Building 12 complex and Building 31. During FY93, we completed the migration of these networks from 3Com 3+Share to Microsoft® LAN Manager and Windows™ NT Advanced Server. In addition, the initial steps were taken to consolidate the administration of DCRT's four production LANs under a single, cohesive organization. Consolidation will result in a significant reduction in the number of production file servers and in the number of people required to maintain those servers.

PUBnet

During FY93, the DSB worked to further develop the NIH Public Network (PUBnet), a network facility designed to provide various network services to users of NIH LANs connected to NIHnet, the campus-wide backbone network. PUBnet consists of a number of network servers that support both PCs and Macintoshes®. Users can link up to these servers from their own workstations and access, download, or run tasks directly from PUBnet. A popular service has been PUBnet's fax gateway, which acts as a hub over NIHnet, giving LAN users at NIH the ability to send electronic faxes from their own network mail to any fax machine, nationally or internationally.

PUBnet's goal is to provide quick and easy access to information and computer tools. PUBnet has achieved this goal in the dissemination of DSB information and tools. During FY93 we included information from other areas of NIH as well. In the coming year, PUBnet will house more information distributed from other ICDs. In light of the vast sea of network-accessible resources around NIH, PUBnet's new goal is to bring together these resources virtually so that users can easily access what they need. Information and tools can be accessed centrally and maintained locally by the organizations responsible. If information is truly power, PUBnet should help NIH to exploit that power.

Computer Security

Computer security is an important concern for users of NIH computer systems. The DCRT philosophy is that proper security procedures are a part of the normal operating knowledge that all who use and manage any computer system need to possess, not a luxury or something to be concerned about only when disaster strikes. In the face of a growing and potentially serious threat posed by computer viruses, the DSB redoubled its efforts to increase the security awareness of personal computer users on the NIH campus. In FY93, a 2-year site license was negotiated and purchased for the F-PROT antiviral program for PCs, which enables it to be used by any NIH employee free of charge. This program is distributed via PCBull, the two User Resource Centers, and via PUBnet, and has proved invaluable in curbing serious virus incidents on campus. Other antiviral, access control, and general security programs for personal computers are currently being evaluated; support for some of them is expected to be announced soon. During FY93, the DSB's Security Coordinator initiated efforts to instruct personal computer users in safe computing practices and promoted the use of DSB-supported antiviral software. Instructional documents and informational bulletins were updated to reflect new viruses appearing on campus. A number of lectures and classes on the subject were also offered. The DSB also developed a plan to establish a Computer Emergency Response Team (CERT) to resolve security incidents across DCRT and to promote security awareness and training.

NIH Scientific Directory and Annual Bibliography Project

In FY93, the DSB led the development of a Macintosh® database and desktop publishing system to produce the *NIH Scientific Directory/Annual Bibliography*, an annual summary of all NIH senior personnel and their publications. In previous years, the SD/AB booklet was produced using the mainframe and arcane embedded Government Printing Office printing codes, leading to many errors and long delays in production. The Macintosh® system greatly simplified the ICD submission requirements in FY93 and was extremely helpful to the Editorial Operations Branch, OD, in proofing and producing this year's SD/AB booklet. Future enhancements based on this year's experience should make the process even more efficient and cost effective in FY94. This project was a collaboration with the DCRT Information Office, and with the Editorial Operations Branch, NIH/OD.

Future Directions

An important part of the move from personal computing to distributed computing is the clarification of the "true cost of computing" to the NIH community. ICDs tend to focus on initial hardware and software costs as being the bulk of their investment in distributed computing. In reality those costs are likely to represent less than 25% of the 5-year costs of using local resources. The rest of the costs come from labor associated with providing support and training, installing software and hardware upgrades, and maintaining administrative control. This represents a shift from central services to local responsibility that is often overlooked when groups establish distributed computing. Without adequate realization of and commitment to these new responsibilities and costs, the quality and productivity of end-user computing is reduced.

Gartner Group, a leading computer industry research and consulting firm, estimates that the life-cycle cost of personal computing has more than doubled since they first calculated it in 1987. Costs have escalated to a degree that approaches the gains in personal productivity realized by computer use.

The available Gartner Group estimate was based on data obtained from the private sector. This year we conducted a study to begin to identify the true costs associated with desktop computing throughout NIH. We contracted with Gartner Group to assist us by providing independent, objective analysis of the data we retrieved via three different phases.

In the first phase, eight managers from various ICDs worked with us to develop and refine a questionnaire, which was subsequently completed by 28 additional NIH managers as phase two of the study. This questionnaire allowed us to obtain data regarding the costs associated with the administration, support, and acquisition of desktop computing. The third phase of the study consisted of a questionnaire mailed to several thousand NIH users (and completed by 700) to obtain data regarding the costs associated with end-user operations. We define end-user operations as those functions related to users learning how to use their systems, developing their own applications (including macros), performing backup, recovery, and file management, and maintaining their systems.

The costs associated with administration, support, acquisition, and end-user operations were compared with the average initial capital investment required to purchase desktop systems.

Next year, as we increase the awareness of NIH users and management regarding the true costs associated with desktop computing, we will try to persuade them to increase support at the local level. We believe administrative and end-user costs can be reduced through greater investment in technical support at the local level. Not only should that support pay for itself, but it will likely promote increased process automation and organizational integration, resulting in more effective use of computing technology on the desktop. The increased complexity of computing on all platforms requires comprehensive support. NIH organizations that put this support in place now will be best positioned to take full advantage of the distributed computing technologies that are rapidly evolving.

Toward our goal of strengthening local support, we revisited in FY93 our Lead User and Macintosh® Support Coordinator distributed support programs. Both programs were refocused to concentrate on ICD personnel whose official duties include local computer support. The structure of Lead User and Macintosh® Support Coordinator meetings was revamped with a new emphasis on sharing of technical information. DSB also introduced a pilot program of intensive technical training. Next year, we will work to merge these support groups into a single group known as the Computer Support Coordinators.

Organizational Consulting

To increase our effectiveness, next year we plan to focus our consulting resources on assistance to organizational groups rather than to individuals.

We will consult with those who make strategic personal computing decisions for their organizations and those who provide first-line support. Assistance will be provided in selecting hardware systems and application software at the organizational level, implementing local area networks, devising cohesive database management strategies, and developing in-house support capabilities. We will put substantial effort into developing and promoting interoperability among the computing platforms at NIH.

In moving toward organizational consulting, more of the responsibility for direct end-user problem resolution will be provided by the newly established Customer Services Branch. The CSB in turn will leverage its support by outsourcing many of the routine requests for assistance. The transition to organizational consulting must be handled very carefully.

Research Projects I: Image Processing and Support of Nuclear Medicine

Multimodality Research Image Processing System

M. Douglas, B.A.
with P. J. Kalkowski, E. Pottala, Ph.D., W. Gandler (DCRT/DSB); R. Levin, Ph.D., G. Sobering, Ph.D.(NCRR); R. Carson, Ph.D. (CC/NMD); J. Frank M.D. (OD/LDRR); T. Zeffiro, M.D. (NIA), A. Polis (NINDS)

The purpose of this project is to develop an image processing system for the study of multidimensional (2D to 10D) data from multiple imaging modalities. This Multimodality Research Image Processing System (MRIPS) is to be based on a common hardware and software environment across NIH and is to be the standard system for macroscopic image processing (PET, MRI, CT, SPECT, echo, etc.) at NIH. Lack of such an environment has impeded imaging research over the past several years. The development of a new system that could incorporate all of the functionality of the many old systems in use at NIH plus new 3D functionality was made possible through funding for the new Laboratory of Diagnostic Radiology Research (LDRR).

DSB staff led the development of the functional specifications for the image processing software and collaborated in the system design. DCRT staff (M. Douglas) and NCRR staff (Dr. R. Levin) serve as co-project officers, supervising development and delivery of the system. The initial product has been delivered. The initial system consists of a central, short-term (2-week) image registry for storage and management of clinical images, format translators for most tomographic modalities at NIH, a preexisting 2D image processing software package, a medical imaging software extension and a 3D image processing software extension. DCRT leads a cooperative testing effort with membership including researchers and systems analysts from nine ICDs. Already more than 75 users are licensed to use the product at NIH. The initial test system is already in use in sites across NIH. Additional modules are being developed under contract, acquired from different vendors and developed by researchers at NIH. These modules can be integrated seamlessly into the system. Integration of one such addition, a morphology module from the Mayo Clinic's AVW Toolkit®, has been completed.

This new platform is the first major common hardware and software environment for image processing for use across NIH. Use of this common hardware and software environment will greatly increase the efficiency and effectiveness of research involving tomographic data by providing advanced image analysis, segmentation, and visualization capabilities, by facilitating access of data from a wide variety of tomographic modalities, and by sharing algorithm and software development.

DCRT will continue to be a primary overseer of the contractors in their development of the image processing software. Over 100 users are expected by the end of FY93. The system will require expansion in order to satisfy the research needs of these users. The first requirement will be the addition of user scripts to make the system perform the routine clinical analyses needed by these researchers. DCRT

plans to assist in the development of these scripts and to assist researchers in the use of MRIPS. More advanced research will require collaborative development of complex modules which will be incorporated in the system. Areas for future development include more comprehensive registration methods, guided and automatic segmentation, volume-of-interest definition, and visualization. DCRT will participate in developing and incorporating advanced image analysis and visualization algorithms into the system. Specific plans include adding 3D volume-of-interest rubbersheet editing, seeded autosurfacing, and the correlation and statistical parametric mapping methods of registration in the coming year and merged volume rendering in the future.

Computer Systems and Applications for Nuclear Medicine

M. Douglas
with P. Kalkowski (DCRT/DSB); S. L. Bacharach, Ph.D., M. V. Green, M.S., N. Freedman, Ph.D. (CC/NMD)

DSB, in collaboration with the Clinical Center's Nuclear Medicine Department (NMD), develops systems for computer-based mathematical analysis, pattern recognition and image processing in support of diagnostic activities in the collaborating institutes. Many applications are directed toward the correlation of function with structure, such as estimation of ventricular function from radionuclide ventriculography or PET scan (functional data) compared to MRI or CT scans (anatomical data). Other applications are directed to special techniques for image-guided surgery. MIRAGe, a general-purpose image processing system, was developed by DCRT and the Nuclear Medicine Department over the past several years to support these applications. Programming was performed by contractors supervised by DCRT and the Nuclear Medicine Department. The completed basic system has been ported to several other NIH computer systems including VAX™ workstations and Macintosh® systems. Many academic and commercial institutions across North America and Europe have requested and received copies of the system. MIRAGe functionality has been included in the design of the new Multimodality Research Image Processing System currently being developed for use across NIH.

There have been six major application areas over the past year:

- A previously developed system for 3D alignment of tomographic cardiac PET emission data using pseudo-attenuation data has been made more efficient, extended to apply to planar images as well as 3D images, and has been integrated into routine procedures in the Nuclear Medicine Department.
- Following alignment of one PET volume to another, the second volume frequently needs to be resampled to permit quantitative comparisons with the first volume. Programs permitting the rapid resampling of volumes based on realignment parameters have been written and incorporated into NMD procedures.
- Alternatively, a region of interest may be defined in the first volume and the contents of this region compared to the contents of the same region within other volumes within a dynamic PET study. Therefore, a program has been written to compute the realignment parameters, transform the region of interest, and apply the transformed region to the new volume. Using this mapping, the program calculates statistics for a given region in each dynamic scan without having to rotate or reslice the dynamic scan.
- Projection algorithms were developed that permit a 3D volume to be projected to a 2D image from any angle. The algorithms allow for substitution of a variety of systems of equations for the correct computation of attenuation of the signal as it passes through the body. These projection algorithms are also a basis for volume ray-tracing algorithms used to render volume data in the Multimodality Research Image Processing System currently being developed.
- A system for precision image-guided surgery is being developed. Conventional surgical procedures are unable to make use of the high-resolution tomographic imaging information for guiding the surgical approach or the resection of the lesion. In

neurosurgery, stereotactic frames are sometimes used to register the preoperative image coordinates with the surgical coordinates. The constraints of stereotactic frames limit their applicability to a small number of surgical procedures, and there has been considerable interest in developing "frameless" techniques for more widespread image-guided surgery. The system being developed collaboratively by the Nuclear Medicine Department and DSB makes use of preoperative images combined with intraoperative images for planning and executing the surgery or therapy, and for correlating surgical findings with information in the images. The system will initially be evaluated in nephron-sparing surgery on patients with Von Hippel-Lindau disease, but the intention is to evaluate the system for other applications, especially to facilitate neurosurgery.

• DSB is also collaborating with NMD in the development of an image processing technique that automatically identifies noninvasive magnetic resonance markers in tagged MRI left ventricular cardiac images. Previous evaluation of heart wall motion has been limited to regional evaluation where the left ventricle is delineated throughout the cardiac cycle and average wall motion assessed. This evaluation is usually accomplished by using simplified models of cardiac shape and motion. Recently, magnetic resonance imaging techniques have been developed that allow the myocardium to be noninvasively "tagged" at end diastole. Specialized sequence of radio frequency and magnetic gradient pulses produce a series of dark stripes in the images due to magnetic saturation along strips of the myocardium. Intersections of two orthogonal sets of stripes produce a grid of noninvasive markers. The motion of these markers can be analyzed as a measurement of contractility of the marked myocardium.

In the future, DSB will continue to develop 3D visualization methods, such as the projection algorithm; multimodality registration, such as the PET cardiac registration method extended to the brain; and fast interactive algorithms for analysis of large volumes of data. Investigation of clinically useful visualization of volumetric data will continue. Simultaneous display of multiple volumes for the visual verification of volumetric alignment will be developed.

Image Management and Communications System (IMACS)

K. M. Kempner
with H. G. Ostrow (DCRT/NSB); T. L. Lewis, M.D. (CC/DIR); E. E. Tucker, M.D. (NHLBI/CB); M. R. Armstrong, M.D., G. P. McMahon (CC/DRD); P. G. Okunieff, M.D., F. Sullivan, M.D. (NCI/ROB); and J. F. Fessler (NCRR/BEIP)

Medical images are an important component of the medical record generated during a patient's hospital stay or clinic visit. Unfortunately, these images represent a difficult-to-manage data source because of the extremely large size of the datasets involved. The NIH Clinical Center (CC), like most university and research hospitals, is attempting to solve the problem of consolidating medical images with the conventional alphanumeric medical record data in the Medical Information System (MIS) to more completely realize the goal of a comprehensive electronic medical record. Toward this end, DCRT, CC, and NCI are collaborating to develop a series of demonstration projects that explore image integration into the electronic medical record. The images of interest range in size from diagnostic electrocardiograms (16 Kbytes) through tomographic scans (256 Kbytes) to conventional film X rays (4 Mbytes).

Standard 12-lead diagnostic electrocardiograms (ECGs) are automatically acquired, interpreted, and stored on magnetic disk using a Hewlett-Packard ECG Data Management System located in the CC. In order to transfer ECG diagnoses and the related waveforms from this minicomputer system to the MIS, a remote ECG workstation is being developed as a serial RS-232 gateway between the two systems. Because ECG waveforms are essentially binary images (black waveforms on white background), and because the number of equivalent black pixels in such an image is extremely low (approximately 0.1%), ECG waveform data are more efficiently stored and transmitted as time-ordered

lists of 10-bit ECG amplitudes, rather than as 2.75K by 3K pixel images.

Chest X rays are routinely obtained within the Diagnostic Radiology Department, and these images are appropriate for integration into the MIS, as well as for transmission to the relevant outpatient clinic where the patient will be seen. In this application, we are currently using a Vision Ten Rita!® system (which contains a gray-scale sheet film digitizer) as an integral part of an image gateway. In addition, we have installed two Rita!®-compatible image display systems. Utilizing the fiber optic network installed within the CC, communication of medical images between the Radiology Department's Film Library and remote sites is now possible. The weekly NHLBI Cardiac Surgical Clinic was the first outpatient clinic to routinely use chest films transmitted over this ethernet pathway.

Future plans include the connection of two General Electric CT scanners into the Vision Ten image transmission and display environment. This will be accomplished via industry standard (ACR-NEMA), ethernet-based communication links to dedicated image servers, which will be added to the teleradiology network. In addition, we are planning a prototype high-speed image communication network based on Asynchronous Transfer Mode (ATM) Switch technology. The ATM Switch will allow 155 Mbit/sec multimedia communications between users. The CT images to be transmitted via this system will be obtained from the CT scanners' dedicated image servers. Custom-designed Radiology Consultation Workstations will be located in the CC's Diagnostic Radiology Department and in NCI's Radiation Oncology Branch. Real-time consultation sessions between a radiologist and a radiation oncologist should facilitate development of radiation therapy treatment plans.

Brain Image Registration

K. M. Kempner
with M. V. Green (CC/NMD); J. F. Fessler (NCRR/BEIP)

The superposition and registration of differing tomographic views is a difficult problem for investigators attempting to correlate brain form (structure), derived from x-ray computed tomography (CT) images, and brain function (metabolism), revealed by nuclear medicine positron emission tomography (PET) images.

Our approach to this problem is based upon a three-stage strategy. First, we are developing practical methods for the accurate and reproducible placement of the head within a tomographic scanner's aperture. Second, we are developing techniques for monitoring head position during the image acquisition process, so that corrections may be made for head movements before the image is generated. Third, we are developing simplified algorithms for the scaling and registration on a digital display subsystem of digitized images from different scanners.

Precise orientation of the subject's skull within the scanner's aperture is monitored and recorded through the use of a Polhemus position/orientation measurement subsystem connected to a PC, allowing simultaneous use of two independent sensors. The development of two inexpensive custom-molded oral appliances allows the Polhemus subsystem's sensor to be fixed to the subject's skull. A novel targeting algorithm was derived to provide visual cues related to head position within a scanner's imaging volume to the system operator. Two-sensor software was completed, and extensive evaluation has begun prior to its experimental use with test subjects.

Future efforts will continue to center on clinical testing of the accuracy and repeatability of skull placement in tomographic scanners, as well as in refinement of algorithms for removing motion artifacts during the scanning process. The position/orientation measurement subsystem will also be interfaced to the Real-Time Gamma Camera Image

Correction System, so that motion artifacts may be eliminated from the small field-of-view gamma camera served by that system.

Real-Time Gamma Camera Image Correction

W. R. Gandler
with K. M. Kempner (DCRT/DSB); M. V. Green, and J. Seidel, Ph.D. (CC/NMD)

The Nuclear Medicine Department (NMD) of the Clinical Center has developed a small field-of-view (FOV) gamma camera which has great promise for practical, high-resolution imaging of small animals. The system is based on a novel position-sensitive photomultiplier tube (PMT), instead of multiple non position-sensitive PMTs used in standard, large FOV gamma cameras. Unfortunately, the position-sensitive PMT does not possess either a linear voltage analog of event position, nor a uniform energy response across the tube face.

Previous collaborative efforts between the NMD and DCRT have, independently, demonstrated the need for methods to correct for motion artifact during planar gamma camera studies of the brain. A technique was developed, utilizing a Polhemus position/orientation measurement subsystem, to perform the necessary corrections.

We are currently developing an Intel® Multibus® II computer system that will allow geometric, energy, and motion corrections to be performed sequentially, in real time on data from the small FOV gamma camera. This image correction system will be interposed between the gamma camera and its data acquisition computer, intercepting and processing the data as they are transmitted from the camera to the computer.

Three coupled 386/486 processors comprise the Multibus® II system. These processors are dedicated to input (analog-to-digital conversion), computation (geometric, energy and motion corrections), and output (digital-to-analog conversion or digital transmission), respectively.

System control software has been developed, as well as programs to acquire data from the analog-to-digital converter modules, and to display both uncorrected and corrected data arrays. All geometric and energy correction software has been completed. Software development is continuing with the completion and testing of output routines for the digital-to-analog converter modules, and to permit the integration of real-time data from an independent PC-based position/orientation measurement system.

Research Projects II: Laboratory Automation and Genomics

Utilization of Specialized Hardware for DNA Sequence Analysis

J. I. Powell

An Applied Biosystems, Inc. (ABI) Inherit™ system was purchased by DCRT and is being made available as a shared resource to the NIH intramural scientists. Scientists can purchase Macintosh® client software from ABI and access the Inherit™ system over the NIH network. Inherit™ makes heavy use of specialized hardware, the Fast Data Finder (FDF), a parallel processor capable of scanning and comparing over 15 million characters per second. This system is primarily designed for: (1) assembly of medium to large sequencing projects, (2) for searching the gene and protein databases for homologous gene sequences, and (3) to quickly search for genetic motifs such as regulatory elements. A pattern search language incorporated in the system allows for very complicated query formulations. DCRT is pursuing the possibility of porting Server software to additional UNIX® platforms, including the NIH Convex.

In connection with the ABI Inherit™ evaluation, a critical, quantitative comparison was done of several different commercial computer programs designed for sequence assembly and analysis. The purpose this study was two fold: (1) to gain experience and expertise in the use of several different sequence assembly programs, and (2) to

evaluate these programs as to their speed and accuracy in assembly and in their ease of use. Six sequence assembly packages have been examined. To evaluate the speed and quality of sequence assembly, the rat multi-drug-resistant gene (RATMDRM, 5254 base pairs) sequence was randomly split into 58 overlapping fragments, each 200-400 base pairs in length. From 0 to 15% error was randomly added to these fragments based on the distribution of error found in the original fragments that were used to find the assembly.

Computational Resources for Automated DNA Sequencing of cDNA Subtraction Libraries Initiative

J. I. Powell

A collaboration is ongoing with Dr. L. Staudt, NCI, on the discovery of novel human lymphoid-specific genes by automated DNA sequencing of cDNA subtraction libraries. Software tools developed by DCRT are used to process and place the data into a SYBASE relational database system. These include tools for prescreening cDNA sequence against a local database, automating searching against the nonredundant databases on the NCBI network BLAST server, providing for the display of the results and allowing user interaction to select information to be placed into the SYBASE database. Work was initiated to provide software to perform complex motif pattern matching analysis on the cDNA sequences.

To date, approximately 1,500 cDNA sequences have been analyzed yielding homologies to a variety of proteins including transcriptional regulators, signal transduction proteins and membrane receptors. Work is in progress to expand the use of the database to include laboratory management information and data from other sources such as northern plots.

Logic Programming-Based Database and Query System for Genomic Data

R. C. Taylor

Computational support was provided for research in comparative DNA sequence analysis and for the construction of public-domain tools for biologists to use in such analysis across multiple genomes. Work continued in the analysis of the global organization of selected genomes and the definition of local regulatory grammars for the genetic regulation of metabolic pathways. A common feature of all such work was the logic programming language PROLOG. Use of PROLOG allowed us to combine data of disparate types and rapidly develop the ability to pose complex queries of the integrated data. Tool-set development was performed in close collaboration with Dr. Ross Overbeek and Dr. Ray Hagstrom of Argonne National Laboratory. The collaboration on biological database construction continues with Argonne National Laboratory, with the current emphasis being on the addition and integration of large volumes of metabolic data to DNA and protein sequence data.

Flow Cytometry Advanced Data Analysis (FC/ADA)

L. K. Barden
with R. Tate, Ph.D. (DCRT/CSL/DSB); S. Sharrow (NCI/DCBDC/EIB); D. Plugge, P. Johnson, J. Robinson (Systex, Inc.)

The Flow Cytometry Advanced Data Analysis project (FC/ADA) is a collaborative laboratory automation project with the Experimental Immunology Branch (EIB), Division of Cancer Biology Diagnosis and Centers, NCI to design and implement a basic research support facility capable of the acquisition, archiving, and in-depth analysis of multiparameter flow cytometry data.

The facility's computer systems permit a number of complementary analytical techniques, such as nonhierarchical cluster analysis and multidimensional gated histogramming, to be applied

to experimental data. Experimental conditions and sample parameters are stored in machine-readable form along with the data.

A data staging and archiving system scaled to match production data acquisition rates is provided as part of this facility. This provides near online access to experimental data for an extended period of time as well as automatic archival storage (and retrieval) of all experimental data.

The EIB flow cytometry laboratory currently supports multiple research projects for more than 40 investigators within EIB and NCI. These investigations involve quantitative single-cell analysis of parameters associated with cell freshly prepared from different species and tissues, as well as a spectrum of *in vitro* cultured cells.

While the direct collaborative effort involves DCRT and EIB, the software and techniques being developed under this project are shared with other flow cytometry facilities within the NIH intramural research program. Flow cytometry sites at NIAID and the FDA Center for Biologics Evaluation and Research are currently involved in this effort.

Current Status and Future Plans

The EIB facility has been operating independently of DCRT for most of FY93. Computer system management and the final phases of system tuning and load balancing are being done by contract personnel retained by EIB. The Cluster Analysis Program (CAP) is being ported from its originally designed VAX/VMS™ minicomputer and graphics terminal environment to a RISC OpenVMS™ workstation platform, with some necessary changes to computational algorithms and data structures to take full advantage of the RISC architecture. We will continue to support and enhance the Cluster Analysis Program under the Open Molecular Analysis Environment being developed by the BioInformatics and Molecular Analysis Section, DSB.

The Hierarchical File Storage System (HFSS) – consisting of a magneto-optical disk jukebox and associated hierarchical file storage software – has completed testing within DCRT and has been deployed to EIB, where it provides 10 gigabytes of near online disk storage for flow cytometry data. EIB expects to expand HFSS storage from the current 10 gigabytes to its maximum of 30 gigabytes during FY94.

The seminar series "Topics in Flow Cytometry" hosted two sessions during the year, with presentations in the areas of computer network connectivity for flow cytometer systems, standardization of flow cytometry facilities and DOS/Windows™ based data analysis programs. The scope of "Topics in Flow Cytometry" will be expanded to include imaging cytometry, with a name change to "Topics in Analytical Cytology."

Modernization of Computerized Laboratory Automation

H. A. Fredrickson
with E. Pottala, Ph.D.(DCRT/DSB); T. Miles, Ph.D., F. Howard, Ph.D. (NIDDK/LBM)

Beginning in 1976, DCRT developed and installed 11 Laboratory Data Acquisition and Control System (LDACS) computer systems for NIDDK scientists throughout Building 2 (now relocated to Building 5). Based on then-current DEC® LSI-11 microcomputers, the LDACS computers were connected to laboratory instruments for control and data collection. Collected data were transferred to a central computer over low-speed serial lines for further processing.

We have replaced the three LDACS computers still in routine use, since commercially equivalent systems were not available, with personal computers. The new systems will perform the same functions as the LDACS but will be connected to the building LAN.

The new computers will be equipped with the appropriate interface components to control the laboratory instruments. Currently available desktop computers are considerably less expensive, offer more performance, and are considerably less difficult to program and maintain than the original LDACS.

The software is modular, offering multiple small routines to perform discrete tasks of minimal size (e.g., temperature measurement) which can be invoked from a general user interface program. The user interface has a high degree of compatibility with the existing LDACS system, and user screens can be easily modified.

It is anticipated that this system will be general enough for use in other research labs at NIH.

High-Speed Diode Array Spectrophotometer

H. A. Fredrickson
with A. R. Schultz, Jr. (DCRT/CSL/DSB); W. Friauf, J. Cole, P. Smith, Ph.D. (NCRR/BEIP); R. W. Hendler, Ph.D. (NHLBI/LCB)

A computer-controlled 100-channel High-Speed Diode Array Spectrophotometer has been developed by the Biomedical Engineering and Instrumentation Program of the National Center for Research Resources and DCRT for the Laboratory of Cell Biology (LCB), NHLBI. It will be used to obtain more complete spectral information about the rapid changes of the reduction and oxidation centers within the enzyme cytochrome oxidase. This enzyme is involved in cellular respiration and is located within the inner lipid bilayer of the mitochondrion.

The electronic hardware consists of two 48-element photodiode arrays, each connected to a discrete analog-to-digital converter and, subsequently, to local storage channels. Each channel is capable of acquiring data every 10 microseconds. A 80486-based personal computer is used to control the spectrophotometer. Timing control signals are transmitted from the PC, and the PC receives data from the A/D channels via a 40-bit parallel interface.

DCRT assisted in developing the PC interface to the spectrophotometer and in creating the data acquisition and control software. The DCRT-developed (now commercial) modeling system MLAB is being used for analysis of the data. The spectrophotometer has been built and delivered to NHLBI, and laboratory testing of both the hardware and the software systems has been completed.

Significance

This instrument should facilitate development of a variety of new laboratory applications. Increased temporal resolution will permit investigation of early events in biochemical reaction kinetics that have been impossible to measure before.

Proposed Course

Additional computer controlled timing for up to five experimental functions such as laser, stop-flow and start analog-to-digital have been added this year. A high-resolution, computer-controlled stop-flow device will be added to the system.

An invention report has been filed by BEIP.

Pulsed Electronic Spin Resonance System

H. A. Fredrickson
with T. J. Pohida, P. D. Smith, Ph.D. (NCRR/BEIP); J. Mitchell, A. Russo, J. Bourg (ROB/NCI)

The objective of this project is to develop a pulsed electronic spin (ESR) apparatus optimized for the study of nitroxides and other compounds of interest under *in vivo* conditions.

ESR is a powerful tool for free radical studies and can be useful in biological work if the problem of excessive attenuation by water can be resolved. The system being developed by BEIP will operate at about one-thirtieth of the standard 9 GHz frequency to alleviate the attenuation problem, and will incorporate pulse techniques developed in NMR, as well as other techniques, to compensate for the thirty-fold loss in sensitivity.

For several years, nitroxides have been a major focus of research in ROB, NCI because of their importance to radiation biology in general and their potential utility for new photodynamic therapy techniques.

DSB provided a DOS-based microcomputer interface to control the high-speed electronics and acquire the data for a fast Fourier transform plot with MATLAB. Initial measurements will be spectroscopic, but imaging will be subsequently undertaken.

Significance

Achievement of the proposed specifications should facilitate research on nitroxides and other compounds of interest in biological systems.

The contribution DBS/DCRT provided to the project is complete. We have developed interface software for BEIP's instrument. The project will continue with BEIP and ROB/NCI.

Research Projects III: Clinical Signal Processing

3D Flow Velocity Reconstruction from Color Doppler Ultrasound Images

D. R. Adam, Ph.D.
with K. M. Kempner (DCRT/DSB), M. A. Vivino (DCRT/CBEL), E. E. Tucker, M.D. (NHLBI/CB), T. J. DeGraba, M.D. (NINDS/SB), and M. Jones, M.D. (NHLBI/IR)

Clinical color Doppler ultrasound technology is a popular, noninvasive, real-time, relatively inexpensive imaging modality, which currently allows the 2D visualization of blood flow within the heart and the vascular system. Doppler ultrasound flow velocity measurement is important for the determination of blood/oxygen supply to various organs, arterial wall shear stress, and blood-tissue gas exchange, as well as for the evaluation of myocardial and valvular function.

When the methodology of Doppler flow measurement was studied, it was found to be in some respects misleading. While commercial systems provide a color display of flow that changes with time, it is actually a simulation of flow velocity. None of the present Doppler ultrasound systems measure the spatial position and orientation of the ultrasound transducer and its relation to the flow direction. The flow velocity values displayed are, therefore, not representative of the velocity along the axis of flow. Thus, the project evolved from the goal of providing a reconstruction and display of the 3D flow profile to include a more basic study of the quantification of the measurement of flow velocity by color Doppler ultrasound mapping.

We have developed a procedure that should lead to an accurate determination of flow velocity. This methodology appears to allow not only a more accurate calculation of velocity profiles, flow volume and resistance, but also better estimates of the pressure drop across valve orifices and stenotic vessels. The calculation of flow velocity near the vessel walls (today filtered out by most systems), should allow estimation of shear stress and evaluation of possible future damage to the endothelial surface. Similarly, quantitative measurement of the velocity profile across artificial cardiac valves may correlate with vulnerability to blood clot formation.

The Doppler flow velocity images are usually displayed in color, superimposed on the gray-scale, cross-sectional structural images of the adjacent tissue. There are several limitations to this technique of flow measurement, some due to the instrumentation and others due to the measurement techniques. This project concentrates on the latter, identifying the main causes of errors and distortion, and outlining the methodology for minimizing them. Our methodology takes into account the spatial position and orientation of both the ultrasound transducer and the vessel being imaged. The ultimate goal is the quantification of vascular flow patterns, thus enhancing the usefulness of this important noninvasive diagnostic tool.

Initially, we have chosen to concentrate on the structure and flow in the carotid artery, due to the simplifications that this geometry allows. Utilizing color Doppler ultrasound technology to image the carotid artery from several positions and orientations, produces a data set capable of generating a 3D reconstruction of this vessel's structure and flow profile.

During this first year of this investigation we have assembled the necessary instrumentation within a clinical echocardiography laboratory to acquire color Doppler ultrasound images along with time-encoded position and orientation data for the handheld transducer. A carotid artery/neck phantom was designed and fabricated to allow for calibration

and testing of both the position/orientation measurement subsystem and the Doppler flow velocity measurement subsystem, in a controlled environment.

We have transferred the flow velocity images acquired from a human volunteer via the HP® SONOS® 1500 ultrasound system into separate digital values of structure and flow velocity onto the Macintosh® Quadra 950™ microcomputer system that is the heart of our image reconstruction system. All algorithms and procedures for correcting the flow velocity readings have been designed and outlined in detail. These include algorithms for calculating the 3D spatial position and orientation (of both the structural and flow velocity values) at a different location in space than the position/orientation measurement device. All software has been described in flowcharts, and many of the routines have been written and debugged.

Future plans include completion of all software and validation of the methodology and the software, using the phantom with known parameters. Also planned are controlled experimental studies, and an evaluation of this methodology in the clinical echocardiography laboratory. Our approach may eventually be adapted by manufacturers of Doppler ultrasound imaging systems for inclusion into such systems. Our approach may eventually find wide use in the noninvasive measurement of blood flow velocity in clinical practice as well as in research.

Diagnostic Electrocardiographic System

K. M. Kempner
with E. E. Tucker, M.D. (NHLBI/CB); and J. F. Fessler (NCRR/BEIP)

The NIH Clinical Center's (CC) heart station uses a computerized system for the analysis of the clinical Electrocardiogram (ECG). Hewlett-Packard's ECG Data Management System (DMS) processes 12-lead ECGs from all patients within the CC. This system collects and processes ECG waveforms, measuring amplitudes, durations and intervals. It also provides a clinical diagnosis, allows editing of the diagnosis after physician review, stores the ECG waveforms and the diagnostic reports, and permits searching the database for patients who meet search criteria.

The medical diagnostic criteria are encoded as IF-THEN production rules contained in a Diagnostic Criteria Set. These rules were written using Hewlett-Packard's Electrocardiogram Criteria Language (ECL) and the Diagnostic Criteria set may be modified by the user to tune existing criteria or to add new criteria.

ECGs are transmitted digitally to the ECG DMS over 2,400 baud dial-up telephone lines within the CC, from many of the 41 ECG machines distributed throughout the facility that are compatible with the ECG DMS. The ECG diagnostic reports, and ultimately the ECG waveforms, will be sent to the Clinical Center's Medical Information System (MIS) for display at any user terminal. A Hewlett-Packard™ ECG Workstation has been installed for use as an RS-232 interface between the ECG DMS and the MIS. Implementation of this bidirectional pathway is currently under way and initial testing of diagnostic report transmission to the MIS has begun. Completion of the communication link for report transmission is expected by early FY94.

General Signal Processing for Physiological and Laboratory Data

E. W. Pottala, Ph.D. and J. J. Bailey, M.D.
with H. A. Fredrickson (DCRT/DSB); E. C. Phoebus (University of Puerto Rico); K. Rasmussen (NICHD/LCE)

This project involves developing and applying desktop and mainframe computer-based processing and analysis techniques to signals produced by devices extracting information in physiological contexts (e.g., ECG, EEG) or by laboratory apparatus (e.g., mass spectrometer).

Many signal processing algorithms can be implemented via the commercially available program MATLAB on various computer systems including: IBM® PCs, Macintoshes®, and the Convex superminicomputer with interconnections via

NIHnet. For continuous physiological signals, a device for converting the analog signals to digital data is required; for other types of (digital) data, additional methods for transferring them in a compatible form to these computer systems will continue to be developed as needed.

Major tasks in this project may involve the development of methods to analyze very large data sets (e.g., up to 10.4 million samples representing 24 hours of continuous ambulatory EGC (AECG) data). These methods include data reduction and/or compression, noise suppression using sophisticated filtering algorithms and/or signal averaging, advanced techniques for pattern recognition, and statistically or mathematically based feature extraction, trend analysis, and construction of spectra where appropriate. Visual inspection of features in the power spectrogram may reveal the essential information. However, in some contexts (e.g. mass spectrometer) the data may consist of a spectrum with overlapping peaks and the objective would be to resolve its principal components.

FY93 Progress

A method adapting Tchebychev polynomials to extract a morphological feature parameter is being evaluated on AECG data. Preliminary results show that this parameter can discriminate between normal beats, supraventricular aberrant beats, and ventricular beats.

Power spectra can be produced by two methods: Fast Fourier transform (FFT) or autoregressive moving average (ARMA) modeling (AR is subset). The stability of power spectra produced by FFT was tested using different sampling rates on AECGs (120 vs 480 samples/s), thereby producing different resolution of R-wave sequences. To obtain the RR variability (RRV) power spectrum, the mean was removed from the data and the data were then multiplied by a Hamming window (length 4096) and then zero padded to 32,768 points for FFT calculations.

Two new algorithms for autoregressive modeling, (i.e. Yule-Walker and Burg methods) have been programmed in MATLAB on the laboratory Macintosh® system. Simulated data were used to verify that the MATLAB algorithms produced the same results as those produced by standard FORTRAN algorithms. AECG data were used to further evaluate the AR models; the spectrograms in controlled breathing studies showed the same features as those produced by the FFT method. Comparative evaluation of the Yule-Walker vs the Burg method was performed using simulated data. The Burg method demonstrated better resolution, with results closer to the actual values determined by the simulation reported in the literature.

MATLAB least squares is being used to resolve principal components of spectrograms from mass spectrometry.

Future Trends

Frequently it is necessary to segment very large temporal data sets into smaller epochs for analysis so that slow trends with the entire dataset can be tracked. However, a major problem with FFT power spectra produced on small datasets (< 240 samples) is the artifact resulting from windowing the data. The effect of these artifacts can be minimized if the power is integrated in a fairly wide frequency bandwidth; however, this may obscure trends in a sequence of such small epochs. The advantage of spectra produced by AR models is the absence of windowing artifacts. Quasi-3D plots of sequential spectra in the literature show that the changes in the peaks of sequential AR produced spectra of small epochs can track trends within the larger temporal dataset. For this purpose the Burg method, which performs better, will be adopted.

Scatterplots of the timing of events within a temporal study with or without a morphological feature parameter can also be used to track trends in very large datasets. These methods can obviously be applied to human AECGs in table-tilt or exercise studies and in simian AECGs where the goal is to track changes in autonomic control of heart rate.

Research Projects IV: Laboratory and Clinical Data Collection and Analysis

The Lipid Analysis Sample Tracking System (LASTS)

R. L. Tate, Ph.D.
with J. Hoeg, M.D., D. Wood (NHLBI/MDB)

LASTS is a comprehensive PC-based system for recording the results of lipid analyses performed on plasma samples submitted to the Molecular Disease Branch (MDB), NHLBI, Lipid Analysis Laboratory. It replaces a manual system that was used by the branch to process data relating to human lipid metabolism disorders in over 7,000 individuals. Identifying information about the samples is entered into databases maintained on a laboratory PC and verified when the sample is acquired. The samples, identified by computer-produced labels, are subdivided for analysis. The system maintains records of the number of samples awaiting each type of analysis, scheduling appropriate test runs when a sufficient number of samples have accumulated.

As each analysis is performed, the results are either captured directly from the analyzer or keyed by the bar-coded label for manual entry. The results to date on each sample are maintained in a database that can be searched by laboratory personnel or the referring physician. Once the validity of the test results has been certified, the sample data are copied to a report dataset that is then transferred to the NIH Central Computer Utility and incorporated into the MDB lipid study databases. Verified data are also maintained locally in a form suitable for access by PC-based database query programs. Statistics about controls and standards are also maintained.

The LASTS system, in production use for over a year, has been modified to accept data from a newly acquired automated analyzer. Authorized users have network access to the results database. The system now locally creates official medical record update forms for NIH Clinical Center medical records, replacing an expensive and time-consuming process that involved transfer of the data to the Central Computer facility for additional processing and printing.

Computer-Assisted Patient Interviewing In Clinical Pharmacy

J. M. DeLeo
with F. Pucino, Pharm.D., K. A. Calis, Pharm.D.
(CC/Pharm. Dept.)

Purpose

Clinical pharmacists are becoming more involved in direct patient care. They dispense medication information; assess medication compliance; screen for adverse clinical events linked to medications; and recognize potential health problems related to drug interactions with foods, allergies, medical conditions, and other drugs. To be fully effective, they must keep abreast of new drug information, allocate more time for direct patient contact, and maintain effective interviewing skills.

A collaborative project between DCRT and the Clinical Center Pharmacy Department was begun in January 1990 to explore potential roles for the computer in assisting with these tasks and skill requirements. The objective has been to develop a computer interviewing system that collects medication histories, dispenses medication information to patients, and detects possible untoward events related to medication regimens, thereby making more pharmacist time available for patients who are not candidates for computer interviewing. Warnings generated by this system could aid in focusing the pharmacist-patient interaction.

Methods

Design criteria have included flexibility in authoring interview scripts, maintaining updated comprehensive online information on medications, and making computer interviewing available in the

NIH Pharmacy outpatient waiting area as well as in a variety of clinical settings supported by the Pharmacy Department.

Progress/Status

Interview scripts were enhanced and supplemented. New sexual and immunization history scripts were added this year.

Completion and integration of all program modules and database components and initiation of formal testing of the completed system have been suspended pending the outcome of CRADA negotiations with the United States Pharmacopoeial (USP) Convention for joint development and distribution of the expanded system.

Our early work has confirmed that drug information database generation is definable as a separable, contractible project.

Significance

Our goal is for the system to uniformly collect and document important clinical information, and produce comprehensive medication history reports without the aid of a clinician. The computer system should be portable, permit direct data entry by patients, and require approximately 40 minutes for a complete interview. Preliminary experience indicates that many patients can enter information accurately with minimal assistance.

In January 1993, the U.S. Congress passed a public law that requires all states that dispense medication under Medicare to collect patient information similar to the information that our system collects. The interpretative computer language developed under this project has a generalized capability for generating an interactive, data collecting interview.

Research Projects V: Statistics and Artificial Neural Networks

Cancer Patient Survival Prediction: A Neural Network Approach

J. M. DeLeo
with G. R. Merlo, Ph.D., C. S. Cropp, M.D. (NCI/DCBDC/LTIB); D. E. Henson, M.D. (NCI/DCPC)

Purpose

New genetic and biological prognostic factor information may significantly enhance cancer patient survival prediction and cancer patient management, and this is relevant to treatment planning. The inclusion of these factors, however becomes increasingly difficult because of combinatoric complexities and the potentially limiting assumptions of existing mathematical methods. Artificial neural networks (ANNs), which have been highly successful in analogous multidimensional pattern classification problems in many engineering applications, may be useful in cancer patient survival prediction and management. These new biologically inspired computing paradigms adaptively learn to classify complex patterns of data; they also permit easy incorporation of new prognostic factors as they are discovered. The purpose of this project has been to explore the potential practicality and usefulness of applying ANN methodology to patient survival prediction in breast cancer and in other anatomical sites. The most basic question addressed in this work is, "Can neural network methodology offer improvement over existing computational methodologies in survival estimation?" In addressing this question, it must always be kept in mind that all methods are eventually limited by the predictive capacity of the actual data employed.

Methods

Methods used include identification of quality databases, establishing appropriate collaborations, developing and testing ANN algorithms, and

comparing results with those produced by more traditional statistical and computational methodologies.

Findings

Most of the new work performed this year has been done in conjunction with the American Joint Committee on Cancer (AJCC) Multiple Prognostic Factors Committee. The two aspects of this work are first, long-range planning, and second, computational methodology exploration using a database for colorectal cancer.

The long-range planning addresses the Committee's objectives of utilizing new prognostic factor information in survival prediction and cancer patient management. The basic recommendations have included differentiating, defining, and assigning the following tasks: database development and maintenance, database management, computational methodology evaluation, cancer site specific organization, and clinical decision support system development.

The computational methodology exploration has been done with a large colorectal cancer database. Various neural network methods are being explored and eventually compared in performance to a Baysian approach, a Cartesian and Regression Tree Classification (CART) approach, and a Cox Regression approach. The ANN models explored include back error propagation, a modified Dystal model, and a cascaded correlation network. The basic data set consists of 14 selected covariates, and we are attempting to predict survival outcome at the end the first 10 years. Proper treatment of censored and missing data are special concerns of the study.

It appears that significant improvement in survival prediction may be achieved with neural network modeling and that we will soon have data from the studies of colorectal cancer to verify this claim. Neural networks represent a kind of problem solving based on highly interconnected, parallel, simple computational nodes that collectively represent virtually parametric-free models. Although similar methods can be found in more traditional statistical approaches, the mathematical descriptions there are usually highly complex in comparison to the connectionist models which have greater intuitive appeal. Furthermore, connectionist modeling is easily inspired by the neurological and cognitive sciences leading to the development of very sophisticated computational models that might not be so easily achieved through more conventional discovery routes. This general robust model building capability is a very important feature of neural network methodology and is an important part of answer to the "improvement" question asked above.

In general, ANN methodology as it has emerged from practical engineering applications has not been subjected to strong statistical oversight. However, it could and should be for medical applications. We could, for example, compute confidence intervals for survival plots, using bootstrapping methods.

Significance

ANNs may have distinct advantages over more traditional computational methodologies for predicting cancer patient survival profiles. Modern molecular biology research continues to unveil new biological, genetic, and molecular markers, factors, and indicators that may be valuable in predicting cancer patient survival profiles. Integrating this new information with traditional clinical pathology information for improved predictions under different treatment regimens should be feasible with an ANN approach.

Receiver Operating Characteristic Methodology Support

J. M. DeLeo
with G. Campbell, Ph.D. (NINDS/BFSB)

Purpose

Receiver Operating Characteristic (ROC) methodology has become well established as an important tool for addressing decisionmaking

uncertainties in medicine and in other disciplines. It evaluates how well a decision strategy classifies retrospective dichotomous (bivalent), or fuzzy (mulitvalued) events, and it provides a rational basis for designing decision strategies that classify prospective events. Prevalence and error cost factors are easily incorporated into ROC-based decision designs. The purpose of this project is to conduct research and development in ROC methodology as applicable to biomedical research, to publicize practical extensions of ROC methodology derived from this research and development, and to provide computational service, support, and guidance in ROC methodology to the NIH intramural research community.

Methods

Project objectives are met by means of a close collaboration between a statistician and a computational methodologist who both have knowledge of ROC methodology and long-term associations with scientific investigators within the NIH intramural research community. This team conducts research and performs experimentation with ROC methodology as it applies to modern biomedical research objectives. Useful findings extending basic ROC methodology are delivered in the form of presentations, published papers, and computer programs for distributed use.

Findings

A software package called ROCLAB has been produced. ROCLAB runs under DOS, and it is user friendly. It computes ROC functions and their useful derived features for discrete and fuzzy class membership data. Decision strategies that account for uncertainties related to prevalence, false classification costs, and fuzzy class membership are easily constructed with ROCLAB.

ROCLAB has been installed in the DCRT Scientific Computing Resource Center (SCRC) in Building 12A. ROCLAB consults are now available by appointment in the SCRC.

Significance

ROC methodology remains an important tool in biomedical research. Enhancing ROC methodology to support biomedical research and distributing ROC computational tools on modern computing platforms are useful services to the NIH biomedical research community.

Statistical Studies

J. D. Malley, Ph.D.

Statistical Inference for Quantum Systems

This highly interdisciplinary project examines the theory and practical feasibility, for biomedical applications, of recently developed statistical procedures that operate on data known to be dominated by quantum mechanical noise. The methodology has been successfully used over the last decade by electrical communications engineers, particularly those involved with quantum optics systems. Currently, the focus on quantum optics problems has yielded experimental results that are a full order of magnitude better than non–quantum methods (e.g., in terms of signal–to–noise ratio). On a practical and theoretical level, these widely verified experimental (non–medical) results show how the statistician, for the first time, has the opportunity to undertake nearly classical statistical decision theory on data that are known, for example, to have no classical joint distribution.

A preliminary, quantum experimental noise analysis has been completed by Hornstein and Shapiro. For possible use of the new quantum statistical methods, attention has focused on PET scans, MRI imaging, laser-driven reduced illumination, direct and confocal microscopy of living cells and tissue, bioluminescent molecular tagging, and enhanced chemiluminescence. While in these areas only marginal gains could be expected, evidently much more is possible in the field of femtosecond spectroscopy. Here it has been found, for example, that use of quantum statistical methods

(i.e., quantum optimal control theory) could plausibly yield a four-fold gain in a key measure of efficiency, when compared with other methods verified experimentally for the study of nuclear motion in a membrane protein.

A 20,000-word invited article on the subject of quantum statistical inference will appear (with Hornstein as coauthor) in *Statistical Science* (November 1993). Moreover, using established theoretical and experimental results, it is now known how the conventional statistical rationales for inference (Bayesian or frequentist) must be sharply constrained when applied to quantum data. Thus, the likelihood principle evidently cannot be used to underwrite Bayes methods for quantum noise dominated experiments, nor can the long–run performance or satisfactory repeatability of experimental inferences be used as a premise for frequentist methods on such experiments.

Algebraic Methods for Data Analysis

This project develops new methods in statistics, both theoretical and applied, using techniques of advanced algebra. Results have been obtained for the general linear mixed model, leading to a simpler and complete determination of all testable hypotheses and, for example, optimal estimates of variance components and improved analysis of data having a structured pattern of correlation.

Examples of such data include the variance component problem for balanced incomplete blocks and the more general partially balanced incomplete block designs. For both models, very simple, cookbook-style equations are obtained so that the researcher can quickly determine the precise form of all optimal unbiased estimates of the model parameters. These designs are known to have optimal experimental design features, making them well-suited for biomedical data analysis with the constraints of limited time and critical resources. For data using repeated measurements on the same case, often one or more data points are missing or were not obtained, either by design or for noncontrollable experimental reasons. Classical methods for analyzing such data usually require that all such cases (e.g., subjects) be dropped from the analysis. This is costly and inefficient. Thus, to satisfy the usual statistical conditions for the standard analysis, it is often required that half or more of all cases be deleted. Moreover, using the reduced dataset can easily lead to spurious findings.

The Expectation–Maximization algorithm has been in use as a broadly successful antidote to missing data, and a restricted version is in the BMDP® statistics package. The method, however, is also known to have convergence problems that are hard to resolve. Convergence may only occur to a local maximum, when it occurs at all, and the standard practice of using different start points for the algorithm introduces another set of mathematical and statistical problems. Using an idea first proposed by Rubin and Szatrowski (*Biometrika*, 1982), we obtained a complete solution to this problem using algebraic methods (technically, Jordan algebras). We now recognize that other well-known statistical methods work precisely because they are special cases of our results. One example is the solution of finding maximum likelihood estimates for stationary time series data. Our algorithm finds estimates for the covariance matrix for the data, even when this variation is known to be constrained by a set of linear restrictions. Using large sample statistical approximations, the researcher can probe for effects in measurements taken over time (e.g., true treatment or grouping variation vs within-subject variation) without having to delete a single case.

Our methods apply to growth curve models, variance components analysis, genetic linkage analysis, time series data, and to longitudinal data often acquired in clinical trials. The monograph, *Statistical Applications of Jordan Algebras*, has now completed two cycles of peer review and been accepted for publication by Springer–Verlag, Inc., in their *Lecture Notes in Statistics* series.

Publications and Presentations

Bacharach S. L., Douglas M. A., Carson R. E., Kalkowski P. J., Freedman N. M. T., Perrone-Filardi P., Bonow R. O. Three-dimensional registration of cardiac PET attenuation scans, *J Nuc Med* 1993; 34(2):311-21.

DeLeo, J. M. A neural network approach to cancer patient survival estimation, Computer Applications for Early Detection and Staging of Cancer Workshop, NCI/AJCC, NIH, Bethesda, MD July 1993.

DeLeo J. M. Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty, Proceedings of the Second International Symposium on Uncertainty Modeling and Analysis, IEEE Computer Society Press, 1993; 318-25.

DeLeo J. M. The receiver operating characteristic function as a tool for uncertainty management in artificial neural network decision-making, Proceedings of the Second International Symposium on Uncertainty Modeling and Analysis, IEEE Computer Society Press, 1993; 141-44.

DeLeo J. M., Pucino F., Calis K. A., Crawford K. W., Dorworth T. E., Gallelli J. Patient-driven computerized medication history, *Am J Hosp Pharm* 1993 (in press).

Gallelli J. F., Pucino F., Calis K. A., DeLeo J. M., Dorworth T. E. A computerized patient-driven medication history reporting system, World Congress of Pharmacy and Pharmacutical Scientists 1993, Federation of International Pharmacy, Tokyo, Japan, September 1993.Graham D. Comparing sequence analysis programs for the Macintosh®, DCRT 1993.

Graham D. Macintosh® options for multiple sequence alignment, DCRT 1993.Graham D. Performing multiple sequence alignments with GCG programs, DCRT 1993.

Graham D. Preparing figures for publication on the Macintosh®, DCRT 1993.

Malley J. D. Statistical applications of Jordan Algebras, Lecture Notes in Statistics, Springer-Verlag 1993 (in press).

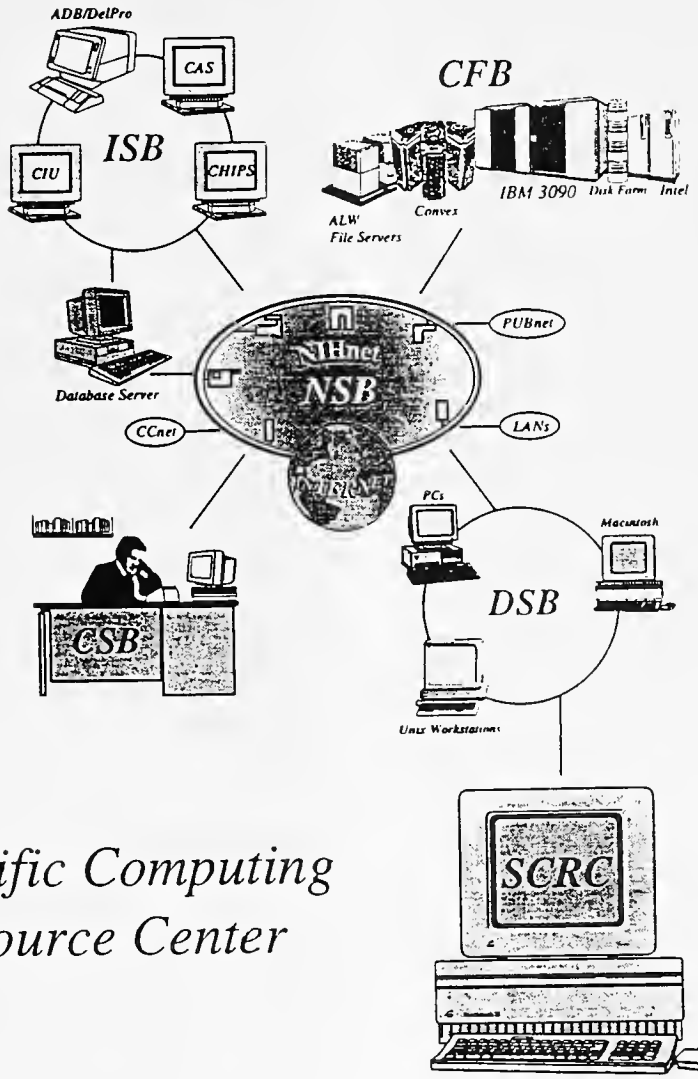
Malley J. D., Hornstein J. Quantum statisical inference, *Stat Sci* 1993 (in press).

Pottala E. W., Bailey J. J., Gilham J. The effect of timing resolution upon RRV spectra with a robust QRS detector after bandpass filtering, *J Electrocard* (Supplement) 1993 (in press).

Zweig M., Campbell G. Receiver operating characteristic (ROC) curves. A fundamental evaluation tool in medicine, Clin Chem 1993; 39:561-77.

SCRC

Scientific Computing Resource Center*



Scientific Computing Resource Center

*The Scientific Computing Resource Center is part of the Distributed Systems Branch. However, at the end of its first year of operation, we choose to give it special recognition in this report.

Scientific Computing Resource Center

Brian McLaughlin, Ph.D., Chief

The Scientific Computing Resource Center, which opened in May 1992, provides NIH with a shared-use computing facility where NIH researchers are able to focus on scientific applications. SCRC staff and other DCRT consultants provide guidance in the selection and effective use of advanced personal computers and UNIX® workstations with emphasis on software for image processing, molecular modeling, sequence analysis, and statistical data analysis.

In addition to advanced PCs, Macintoshes®, and UNIX® workstations, the SCRC facilities include high-resolution image acquisition tools (scanners, digital cameras, film scanners), graphics workstations tailored for molecular modeling applications, and high-quality output devices.

The focus of the SCRC is on the evaluation of software for the analysis of scientific data. Any NIH employee may use the scientific applications in the center for evaluation purposes or for an occasional short–term project. A major goal of the SCRC is to make available different types of scientific computing solutions, so that researchers can make informed decisions about which resources are most needed in their laboratory or office.

The First Year in Operation

During this last year, 735 clients used the SCRC facilities on more than 2,000 occasions. The most popular application areas have included molecular modeling, statistical analysis, image processing, and sequence analysis. The high-demand computer platforms have included the Macintosh® Quadra 950™, the Silicon Graphics® Indigo®, and the COMPAQ® SYSTEMPRO® 486.

Researchers and their associates from almost all of the NIH institutes, divisions, and centers have used the consulting services and resources available through the SCRC. Researchers in three institutes (NCI, NIDDK, and NIMH) used the SCRC the most with regard to the total number of clients, total contacts and total hours of usage. However, on a *per capita* basis, the heaviest users also included NCRR, NICHD, NEI, and NIDR. The SCRC is currently used by an average of 40 researchers per week, mainly by appointment, with an average of 50 new clients each month. The demand for services is increasing.

Application Area Highlights

Image Processing

The SCRC Image Technology Center (SCRC-ITC), which opened in July 1993, provides a variety of image acquisition and processing capabilities. In the first month alone, 14 researchers completed more than 63 hours of scientific imaging projects using the SCRC. The highest user demand in the ITC is for consulting services using the NIH IMAGE program developed by NIMH's Wayne Rasband. The more popular application areas include electrophoretic gel scanning, area measurements, spatial comparison, and image enhancement. Increased usage is expected for image processing resources when the COMPAQ® PC with Imagepro™ and the Hewlett-Packard™ workstation running the Multimodality Research Image Processing System are phased in.

Molecular Modeling

The SCRC provides access to a variety of molecular modeling software, including three of the most popular multipurpose molecular modeling programs: Quanta®, Sybyl, and Insight II®. The software is operational on two Silicon Graphics® Indigo® workstations equipped with high performance graphics capabilities. The molecular modeling software available through the SCRC can be used to assist with the study of a wide range of biological molecules, including proteins, peptides, nucleic acids, polysaccharides and organic compounds. Applications include molecular structure

prediction, protein structure-function relationships and drug design.

Sequence Analysis

Because no single sequence analysis software adequately addresses every research need, a number of different sequence analysis programs have been installed in the SCRC. Their emphases range from determination of primers for Polymerase Chain Reaction (PCR) and sequencing reactions, to software for reading sequencing gels, to high-end programs designed to cover many major analytical needs in the laboratory.

Statistical Analysis

Software packages for statistical data analysis and graphics (both presentation and analytical) are available in the SCRC. These programs, installed on several hardware platforms, support a range of statistical analyses including regression, analysis of variance and covariance, categorical data analysis, general linear models, nonlinear curve fitting and 3D modeling. The SCRC also pursues an NIH-wide campus statistical service through:

- A networked group of ICD clinical researchers and mathematical statisticians, to assist scientists with statistical data analyses and data management problems
- An ongoing series of lectures, workshops and short courses on selected topics. The topics chosen are those that have been shown, by user feedback and recurrence in consulting, to be of the greatest value to the research community.

The Future

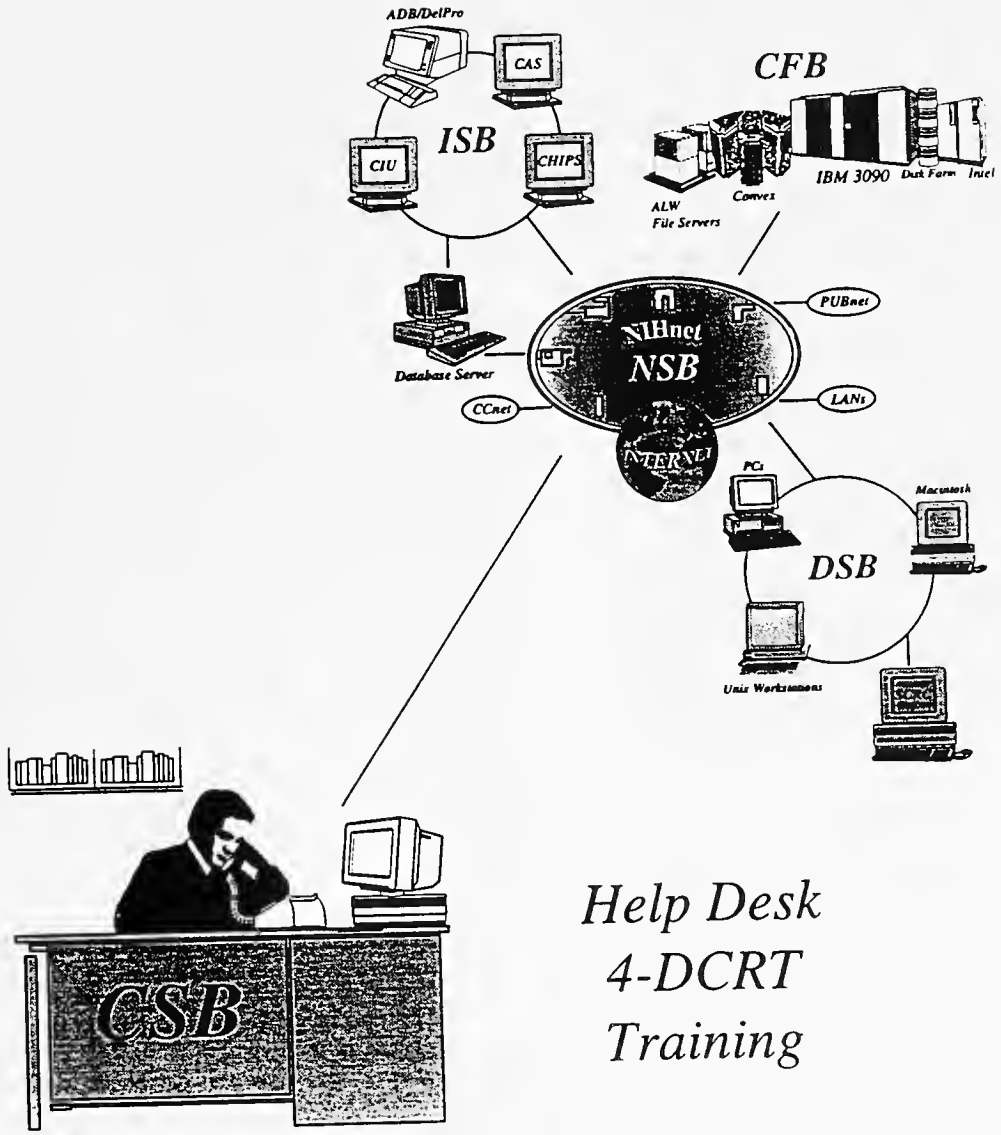
After the end of the first year of operation, a strategic planning and review process was initiated. The pilot phase of the SCRC was evaluated, and this process provided direction and focus to the SCRC. It is important for the SCRC to continually readjust and realign its scope and directives to reflect the strategic plans of DCRT and NIH, and in response to feedback from our staff and clientele.

Over the next 12 to 24 months, other areas of activity will be explored. These will include additional tools for genomic research, enhanced technical graphics support services, resources for the integration of video and sound into scientific computing, and network access for SCRC applications.

Frequently, scientists arrive in groups of three or more in order to examine SCRC software or consult with SCRC staff. We hope to be able to add a combination demonstration and work area that is accessible to groups of three or more users simultaneously to accommodate this requirement.

Additional services under consideration include an information resource library, coordination of beta-testing and "seed" programs for new software products, expanded hours of operation, and the coordination of electronic user forums for SCRC-related information.

CSB
Customer Services Branch



Help Desk
4-DCRT
Training

Customer Services Branch

Dale R. Spangenberg, Chief

The Customer Services Branch (CSB), established during the second half of FY93, is the newest branch in DCRT. A keystone in the DCRT reorganization, the CSB centralizes all of DCRT's initial customer contacts for services and support. When fully operational in FY94, CSB will provide integrated support services in three areas: help-desk consulting, training, and technical information. As the primary liaison between DCRT and its customers, CSB advocates user needs to DCRT management and represents DCRT's expertise, services and policies to its customers.

CSB was created primarily through the realignment of existing DCRT personnel and functions. During FY93, the four other service branches in DCRT identified individuals and activities to transfer to CSB. When fully staffed, CSB will have approximately 20 employees, including computer specialists, computer assistants and a writer-editor. In addition to having career goals in customer support, CSB staff members will either develop or have broad knowledge of DCRT resources, and expertise across various computing platforms. It is expected that the centralization of support services within DCRT will improve DCRT's ability to respond efficiently to the increasingly complex, multiplatform support needs of NIH.

Help-Desk Services

CSB's help desk will provide "first response" to the many questions and service requests received at DCRT. A knowledgeable staff, trained in help-desk techniques, will answer most questions. A sophisticated tracking system will provide status on individual calls and serve as a knowledge base of DCRT expertise. Eight to 10 help-desk specialists will respond to the estimated 400 calls per day that CSB will handle. For those questions that are beyond the expertise of the CSB staff, the caller will be referred to the appropriate resource within DCRT or other support resource. The CSB help desk will reduce the amount of time other specialists throughout DCRT spend responding to customer problems and allow them to concentrate more fully in their technical specialties.

Training

DCRT has established a successful training program in all areas of DCRT expertise and services. This function, headed by Leslie Barden, was transferred from the Computer Facilities Branch to CSB in July. The training function will maintain its high level of service and looks forward to benefiting from integration with DCRT-wide help desk and technical information activities to maintain a keen awareness of customer training needs.

Technical Information

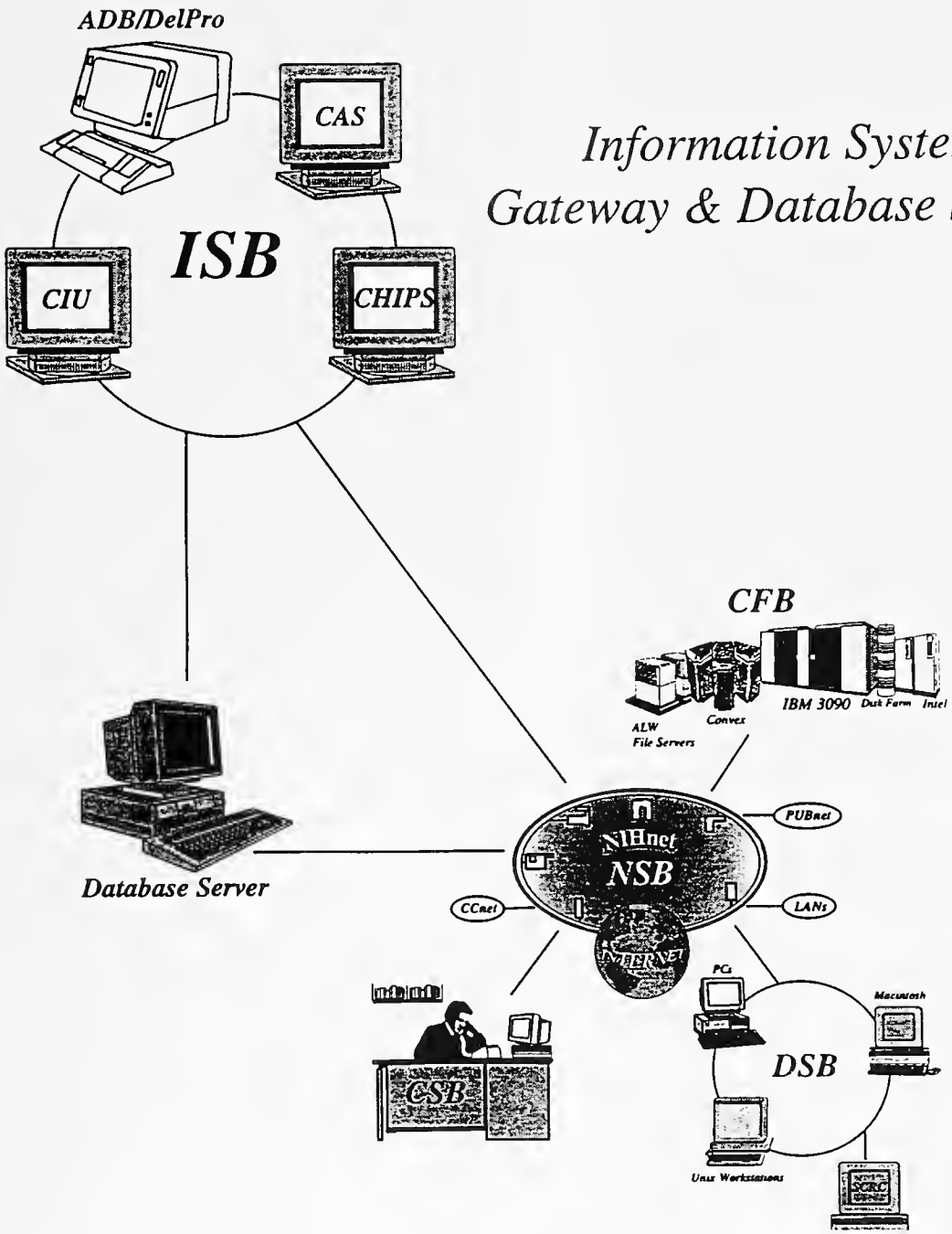
One of the easiest and most cost-effective ways to improve computing productivity is through the effective dissemination of technical information. This function, planned for transition to CSB in FY94, will include the Technical Information Office, currently in CFB, and other information dissemination functions. A variety of printed and electronic forms of information dissemination are planned. This function will sponsor online information beneficial to the user community and help-desk staff in resolving computer-related problems. We are also examining the potential advantages of producing a single DCRT newsletter that covers all computing resources and services.

Future Expectations

CSB is looking to the future with anticipation. The concept of integrated support services for all of DCRT presents exciting opportunities to provide excellence in customer service and to showcase DCRT expertise. CSB will become the central repository of knowledge about NIH computing activities and make that information easily

accessible through a variety of means. CSB plans to augment its staff resources with a substantial complement of technology-based support resources, facilitate cross-training of DCRT staff, and coordinate various user-group activities to provide a healthy exchange of computer-related information throughout the NIH community.

ISB
Information Systems Branch



Information Systems
Gateway & Database Servers

Information Systems Branch

Marvin Katz, Acting Chief

Although its name has changed from the Data Management Branch to the Information Systems Branch (ISB), reflecting more accurately its function, the branch continues to be a central NIH resource which provides advice and services to the NIH user community in the development and maintenance of computer based information systems. The ISB provides advice and assistance to research investigators, program officials and administrators throughout NIH in planning for and obtaining computer information services. The branch also develops, maintains, and processes the NIH Administrative Database and the Clinical Center's Clinical Information Utility. On the staff are 43 permanent full-time employees whose disciplines include computer science, mathematics, and information systems.

The branch is composed of four sections:

- The *Applied Systems Programming Section* (ASPS) provides general analysis, design, and programming services to the NIH community.
- The *New Technology Analysis Section* (NTAS) provides advice on the results of evaluations and proper use of selected new technologies. The section is responsible for analyzing and selecting new database management approaches and for developing the techniques which will facilitate their use across multiple platforms.
- The *Data Base Applications Section* (DBAS) has as its major responsibility the development and maintenance of the Administrative Data Base System, which provides broad support for all administrative and financial processes at the NIH.
- The newest section, the *Data Base Information Section* (DBIS) is responsible for providing the NIH user community with information stored in the Administrative Data Base System. This takes the form of both batch reporting as well as online *ad hoc* queries using graphical user interface, client/server, and relational database technologies.

The NIH Administrative Data Base Supports the NIH Mission

The NIH Administrative Data Base (ADB) represents a major effort by the NIH to combine the administrative and financial data of its intramural program. Using an integrated approach to database management, the ADB concentrates on the full sharing of data among all subsystems that support the NIH intramural program. It features online point-of-origin data entry, minimized data redundancy, background generation of all accounting transactions, and fully synchronized information processing, i.e., the user will always obtain the latest state of any process, data, or function.

The development of the ADB is an ongoing project that encompasses the following features:

- the purchasing, receiving and payment of goods and services is fully supported
- items in nine inventories are individually tracked and are made available by way of online stock requisitions and are completely integrated with the operation of the self-service stores
- all vendors, vendor credits and vendor source agreements are maintained and tracked
- NIH cashier functions are fully supported
- sixteen service and supply fund activities have been or are being integrated
- domestic and local travel orders and travel vouchers are processed and tracked, and foreign travel is being pilot tested; Clinical Center Patient Travel is also supported
- an AIDS Loan Repayment System, to support the repayment of outstanding student loans for scientists who are conducting AIDS-related research at NIH, is integrated into the ADB and utilizes its procurement, invoice and accounts payable functions
- an NIH-wide property management system is functional and data capture is initiated by the receiving module
- the implementation of full research contracts support is under development, and accounting

functions such as fund formulation and funds certification have been shifted to online ADB support.

A more specific summary of new ADB initiatives during FY93 is presented below:

• *Request for Purchase Action (RPA).* The RPA provides ICD laboratory/branch personnel a facility to enter a purchase request "worksheet" into the ADB and to electronically post it to the administrative office for processing. The software to support this function was completed during FY93 and was pilot tested by Telecommunications Branch and Printing and Reproduction Branch personnel. The RPA capability is currently being phased in throughout the NIH.

• *Service and Supply Fund Activity System (SSFAS).* Printing and Reproduction Branch services were implemented during FY93. This function includes ICD service request data entry, establishment of Universal requests, tracking of work requests, billing, interfacing with the Division of Research Grants (DRG) for printing of grants, and council books, interfacing with copy center copiers and processing of Government Printing Office bills for services.

• *Travel.* Domestic, foreign, local and patient travel advance and voucher processing are all currently in place and their usage is mandatory for all ICDs. Sponsored Travel (also known as "348 travel" or travel "in-cash/in-kind") is currently under development and will handle the establishment of receivable entries in the Central Accounting System for collection of sponsor commitments to NIH travelers.

• *Property Management.* Phase II of the Property Management System which supports property passes and the printing of personal appeal forms, was completed during FY93. We are currently in the process of completing the NIH annual inventory and reconciliation for all property items in the property system.

• *Inventory Management (Self-Service Stores).* A new Self-Service Store was established within the ADB to support the NIH staff that currently is located at Executive Plaza in Rockville.

• *Radioactive Materials Ordering System.* The Radioactive Materials Ordering System, which will operate under the Administrative Data Base, will allow the Radiation Safety Branch (RSB) to consolidate on a daily basis orders from the ICDs for radioactive materials and release the summary order to the vendor. Currently, RSB processes about 100 orders daily. Each order must then be individually received and invoiced. Through the proposed system, that number would be reduced to only three orders a day that need be received and invoiced! In a related benefit, there would be a controlled delivery point for the radioactive materials with the potential for future inventory control and distribution point for high-use items. The ICD-entered NIH88 radioactive materials control data would electronically interface with the RSB VAX™ machine, thereby eliminating the need to manually match this form with each ICD radiation materials order.

• *Decentralization of Procurement Functions.* During FY93, the National Cancer Institute was given certain procurement authorities to process a majority of their own procurements without going through NIH Central Procurement. Currently we are working with the Clinical Center (CC) and the National Institute of Diabetes and Digestive and Kidney Diseases to establish the same authorities within their organizations.

• *New ICDs.* During FY93 three new ICDs – NIAAA, NIDA, and NIMH – were included as part of the ADB user community. This required a significant effort in terms of setup, training, and transition from previous systems.

• *Financial Management.* Currently we are working with the Budget Execution and Financial Reports Branch, DFM to facilitate an online mechanism for collecting spending plans by Budget and Sub-Budget Activity and tracking allowance spending against these plans.

• *ADB Security.* The ADB user passwords were expanded from three characters to six characters variable length during the year to satisfy the Chief Financial Officer (CFO) auditors' recommendation.

• *Fellowship Payroll.* During FY93, we incorporated the Automated Clearing House (ACH) function to

handle Electronic Funds Transfers (EFT) for monthly fellowship payments. This effort has established the groundwork for utilizing this procedure to pay NIH vendors for goods and services as well as to reimburse NIH employees.

Contract staff is maintaining the software and documentation for the procurement, inventory, accounts payable, and support subsystems. During the fiscal year, the maintenance staff completed and placed into production 40 change requests.

Administrative Data Base Information System (ADBIS) Provides Easy Access To ADB Information

In collaboration with over 70 ICD representatives, ISB staff has developed a pilot information system which will provide timely and accurate information from the ADB to the NIH user community. This effort involves identifying user needs, developing prototypes and implementing those that are useful management information solutions. The long-term direction for this effort is the eventual development of an Executive Information System in the ADBIS. The ADBIS uses Graphical User Interface (GUI), client/server, and relational database technologies to make such information readily available to the end users on their workstations.

The ADBIS was introduced NIH wide in July 1993. Four functions were made available at that time: FY Obligation and Commitments, Service and Supply Fund Activities, Stock Requisitions and Procurement. The Market Requisition information was made available in early August, to be followed by Travel in the Fall. The Property Information and other ADB systems will follow.

The functions available in July have been tested by the NIH Information Committee's pilot user group during FY93. Presentations of the ADBIS have been given to the ADB Steering Committee, the NIH Intramural Administrative Officers and the NIH community in general.

Other ISB-Supported Computer Applications Highlights

Clinical Information Utility

Developed during the 1970s as an historical archive of clinical information for research, the Clinical Information Utility (CIU) gathers data from the Medical Information System (MIS), the Medical Records Department, and the various service organizations in the Clinical Center (CC). Over the years, millions of records have been archived and made available for use in ongoing research protocols, and for retrospective search and display. The CC Information Systems Department monitors and authorizes all users of CIU data, and the CIU automatically tracks and reports each access of the database. To satisfy clinical investigator needs, the CIU currently handles approximately 10 recurring and 20 *ad hoc* requests each week.

The CIU continues to assist the Medical Record Department in the creation of a series of studies to track the amount of time patients spend in the hospital for various diseases. Ongoing studies in women's health research will require the development of additional programs by the CIU.

The CIU is working with the Medical Record Committee to establish procedures for the presentation of historical laboratory results in System International Units (SIU) in lieu of the current standard lab values. Reports are being developed that will be available for any researcher requesting these data. The information contained in these reports will be the current units, conversion factor, reference intervals and the calculated SIU values.

The CIU is also working with the Medical Record Department to develop a method to determine the accuracy of data entered in their department. Procedures will be developed to compare Discharge Diagnoses data with data from a Discharge Analysis Register to determine whether data have been entered incorrectly or not entered at all.

Child Health Information Portfolio System

The Child Health Information System (CHIPS) provides a central facility for timely and easy access to the Information for Management, Planning, Analysis, and Coordination (IMPAC) system and NICHD-specific data for grants, pending applications, jointly funded awards, and subproject and intramural grants for current and all past fiscal years. CHIPS assists NICHD staff by providing tools for the analysis and management of research grants data.

During FY93 a procedure was developed to automatically transfer scientific and administrative program assignment information from the initial application to all future fiscal years of support over the life of a grant. As a result, assignment data are available for analysis and reporting as soon as a grant for a future year is identified in the IMPAC system, and program assignments can be transferred back to DRG for updating IMPAC within 24 hours. Over 2,100 grant assignments were made in a period of 4 months, resulting in a tremendous time savings for NICHD.

Collaborative Project with National Institutes on Alcohol Abuse and Alcoholism

A collaborative effort with the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and ISB staff has resulted in the successful evaluation and selection of the technologies to be used in the NIAAA Clinical and Research System. This effort utilized the expertise of ISB staff in database, client/server, and related areas to help select the best architecture for the new NIAAA system. NIAAA will continue to consult with ISB staff for future stages of the new NIAAA system development.

System Modeling

A team of ISB systems analysts has been investigating the use of information modeling as a more structured and less technology-centered approach to systems application development. The mission of the system modeling team was to become educated in the methodologies, techniques, and tools supporting model-driven application development. Using data, process, and logic modeling techniques, the team performed the analysis for a pilot project, Request for Purchase Action (RPA). RPA, a current subsystem of the Administrative Data Base, provides ICD staff the capability of electronically submitting requests for goods and services to their respective ICD ordering office. The Bachman Analyst CASE tool was used by the team to implement the modeling techniques and automate the graphical diagramming processes. The major goals of the modeling effort were 1) to establish a well-documented methodology which could perhaps serve as a standard for future development efforts within the branch; 2) to generate documentation that would be appropriate for confirming project requirements with an end user, and 3) to produce a programming specification document that could be passed to a programmer to be manually implemented in the target language of choice. A document detailing the experiences and findings of the system modeling group is currently being prepared. A formal presentation has also been planned for later this year.

Collaborative Effort with NIH/OD/Executive Secretariat

As reported last year, a collaborative effort is ongoing with the NIH/OD/Executive Secretariat to upgrade a correspondence tracking system which is a multiuser, dBASE III® system running on a local area network (LAN). Several client/server strategies were prototyped. We had demonstrated the feasibility of running the system as a dBASE IV® front-end to the Microsoft® SQL Server, but experienced problems inherent in the open-systems concept where hardware and software from various vendors must interface. We have worked with the vendors to resolve these problems and, as client/server technology has matured over the past year, have upgraded our prototype system to include new software versions. We now feel confident in this

technology and plan to implement it during the next 6 months. Implementation of the upgraded correspondence tracking system awaits OD's receipt of all necessary hardware and software. Once installed, the old and new Executive Secretariat systems will run in parallel until it is deemed safe to switch all processing to the upgraded system. To further help the OD, we are currently assisting them in acquiring optical disk technology for the storage and retrieval of documents.

Clinical Center Medical Record Department

Collaborating with the CC's Medical Record Department, the New Technology Analysis Section developed a second prototype system in dBASE IV®, one of the DBMS packages selected for evaluation of client/server technology. The Computerized Microfilm Index System (CMI) is an advanced information system that establishes a relational database and performs the functions necessary to track the process of microfilming inactive and multivolume patient records as well as the reactivation and flowback of medical records. The system will provide a single point of entry for the many specialists involved in the tracking process and will coordinate their efforts into a smooth flowing process. Processing is currently being performed manually and separately without a common database. To further evaluate client/server technology, a second system, the Medical Record Charge-out System, was developed to automate the record charge-out process of retired records. Although the CMI and the Medical Record Charge-out System are two separate systems, they are brought together and presented to the user as a single, total system using relational database and client/server technology. The Medical Record Department is establishing their hardware and installing software in preparation for system implementation.

A significant challenge in the evaluation of client/server technology is the integration of multivendor software into a total system. The above prototypes successfully demonstrate the feasibility of using a mixed product line to develop a system that provides a graphical user interface front-end to a SQL database on a local area network. A successful client/server strategy must also include tools for quick and easy access to the database from a variety of platforms. Next, we plan to evaluate client tools for *ad hoc* querying and reporting. The Executive Secretariat and CMI prototype databases will be used in this evaluation. Multiplatform tools that run on both PC and Macintosh® workstations will be explored.

PC-based Application Development Tools

The Easel Workbench® product is being evaluated to determine if it is a viable graphical user interface development tool for use by ISB system developers. The evaluation involves developing a GUI for the Request for Purchase Action feature of the ADB, which is currently using a character-based user interface. The Easel Workbench® product is being evaluated as the application development tool. The Easel/Win Production System component of the Easel product line is being used to provide run-time execution services for the Windows™ platform.

Other Projects

The ISB continues to collaborate with the NCI Laboratory of Pathology (LP) and the CC to maintain and enhance the NIH Pathology Language Encoding System and the NIH Pathology Retrieval System. Since October 1992 these programs and associated linguistic and semantic dictionaries and rule systems are used weekly to process the LP surgical pathology reports for input to a database maintained for LP within the Clinical Information Utility (CIU). This database is then searched for resp onse to queries Dr. Elaine Jaffe, LP, LP staff members, and others by their permission. Retrieval is by the diagnostic subject matter of the reports with a high or low degree of specificity as required for research purposes. The Pathology Language Encoding System runs on the IBM® 370. While

queries for the Pathology Retrieval System are generated on the Convex, they are executed on the IBM® 370.

The Ethics Information System, developed by ISB Applied Systems Programming Section and sponsored by the NIH Division of Personnel Management, provides a central facility for the collection and review of information utilized in tracking conflict of interest. Using images of preprinted DHHS and NIH forms, data are collected, modified and reviewed through a forms processing software package (DVision®) operating on the OD LAN. Authorization and approval is granted by use of electronic signatures. The forms data are further queried through a dBASE IV® application on the OD LAN. Data entry and review of personal records is available to all ICD employees accessing the OD LAN. The automated access to combined ICD's data through the DVision® forms process and the dBASE IV® DBMS is restricted to authorized users. Currently the available forms include data collected for authorizing outside activities, sponsored travel and guest worker assignments. A WYLBUR command procedure is available to query the mainframe database of the Division of Research Grants' grants and contracts. This information source is also included in the determination of conflict of interest. The pilot phase will include NCI's Division of Cancer Epidemiology.

The ISB Applied Systems Programming Section continued to provide ongoing support, analysis, design and maintenance for new and long-standing projects during the FY93. Included among these are the following:

- for the Fogarty International Center, continued support of the Visiting International Scientist in America Management Information System
- for NIMH, NIAID, and OD, continued support of the Full Time Equivalency Management System
- for DRG, continued support of the NIH Consultant File which is used to assist NIH staff in identifying potential members of NIH advisory committees
- for NCI, ongoing support of the Grants Literature System and Grants Elemental Network Internal User System
- for the CC, continuing support of the Human Leukocytes Antigens Donor System used by the Department of Transfusion Medicine
- for NCI, development of a database system for the Chemoprevention of Prostate Cancer project
- for the NIH Transportation Branch, development of a database tracking system to record maintenance on all NIH vehicles
- for the Division of Nutrition Research, continued support of the Human Nutrition Research Information Management System.

Future Plans and Challenges

During FY94, ISB plans to devote a considerable amount of time in responding to ADB security, control, and documentation issues that were presented in the CFO audit report.

To further expand the ADB reporting capability, we plan to convert Property and Travel data stored in the ADB to the relational ADB Information System (ADBIS).

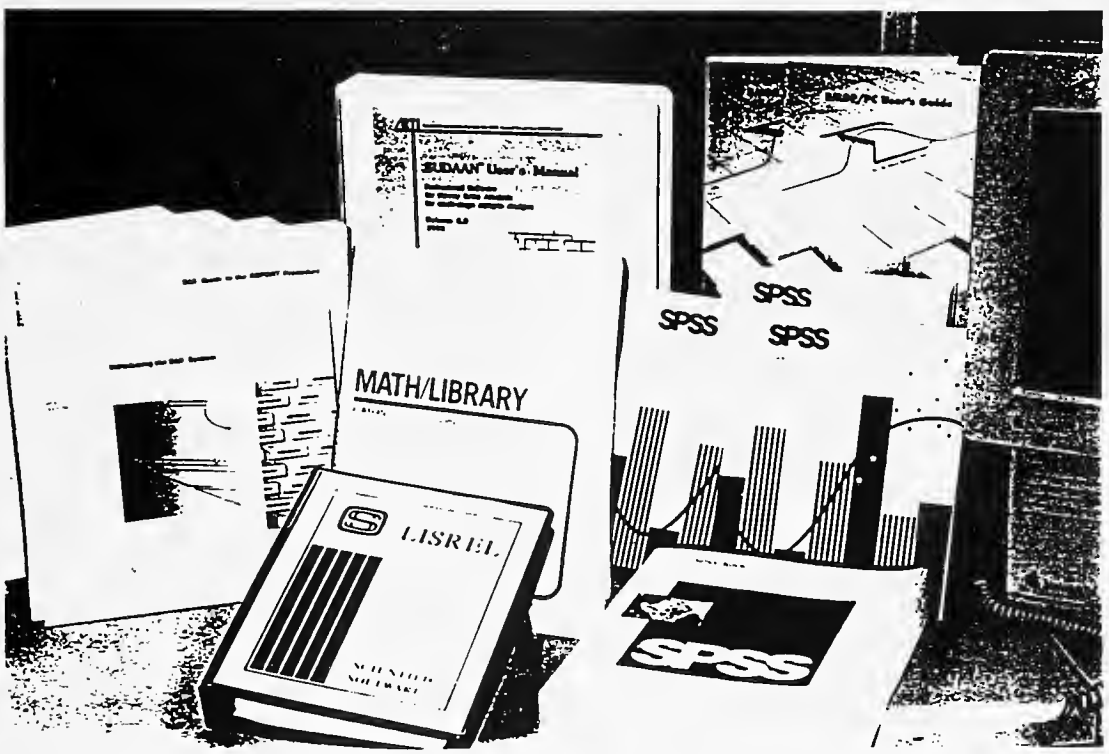
A client/server graphical user interface prototype has been developed by ISB staff to demonstrate the capabilities and features of these new technologies to access data in the ADBIS. In the future, we hope to develop an alternative front end to the ADBIS using these new technologies.

As information systems professionals we are further developing our skills and expertise in structured systems analysis and design. In so doing we will be better able to serve the NIH in helping end users define their requirements. We will also be providing support and consultation for the end user in the planning and analysis phases of the system development life cycle. The final product would be a well-documented system, including all requirements and ready for the construction phase, i.e., programming and implementation. Based upon our work in the system modeling project, we are further refining a formal methodology in how information systems will be developed and maintained. Closely related to ISB's overall strategic plan (which complements the DCRT strategic plan) is the challenge to re-engineer

our legacy systems, i.e., our existing systems, in particular, the Administrative Data Base and the Central Accounting System. In re-engineering, our challenge is to capture the functionality of the current system through a reverse engineering process and then forward engineer the system applying new technologies. Through this approach we hope to leverage the investment that has been made in these systems over the years. We will work on the preparation of the NIH ADB for migration to new technologies in the next 2 to 3 years.

OAD, OCRS

Office of the Associate Director, OCRS



Office of the Associate Director, OCRS

J. Emmett Ward, Acting Associate Director

The Office of the Associate Director, OCRS, consists of the Statistical Support Staff (SSS) and recently created Architectural Management and Funding Management Staffs. SSS is described below.

Statistical Support Staff

Ray Danner, Chief

The Statistical Support Staff (SSS) is in the Office of Computing Resources and Services. The SSS, comprised of six individuals with mathematical and statistical backgrounds, is responsible for providing NIH scientists and administrators with a wide range of services concerned with the application of computer technology essential to NIH programs. This group provides: (1) a combination of research in mathematical statistics and computer information science with collaboration and service in all computational aspects of biomedical data analysis, (2) advice and consultation on the quantitative analysis of biomedical research data and use of the computer in such analysis, including interpreting output and developing statistical procedures when needed, and (3) selection, maintenance and support of standard mathematical/ statistical software for general use of research investigators and administrators in the NIH community. Support includes training, advice and assistance on the proper use of the available software.

SSS provides statistical, mathematical, and other scientific systems and packages to the NIH user community, and evaluates new systems and packages for suitability to NIH needs. Computer systems and packages supported by SSS are shown in Table 3.

Use of mainframe statistical packages at NIH remained at a high level in FY93.

As in previous years, the SAS® statistical and

Table 3. Systems and Packages Supported by SSS

SAS, SAS/GRAPH, SAS/ETS, SAS/OR, SAS/FSP, SAS/AF SAS/IML, SAS/CBT101, SAS/CBT102, SAS/CBT106, SAS/INSIGHT, S/QC, AS/CALC, SAS/TOOLKIT, SAS/ASSIST, SAS/DB2, SAS/CONNECT, SAS/STAT
Vendor: SAS Institute, Inc. A batch and interactive IBM S/370 system for statistical analysis, with extensive file manipulation and graphics capabilities; also in interactive mode on MS-DOS machines.

RPART
Public domain SAS procedure which performs the recursive partitioning analysis routines of J. H. Friedman.

SPSS, SPSS/TABLES, SPSS-PC+
Vendor: SPSS, Inc. A system for univariate and multivariate statistical analysis with file handling capabilities; supported in batch mode on the IBM S/370, and interactive mode on IBM S/370 and MS-DOS machines.

BMDP
Vendor: BMDP Statistical Software, Inc. A collection of IBM S/370 batch programs for univariate and multivariate statistical analysis.

IMSL (International Mathematical and Statistical Libraries)
Vendor: Visual Numerics, Inc. An extensive collection of FORTRAN routines for statistical and mathematical analysis; supported for IBM S/370, Convex and MS-DOS machines.

SUDANN, SESUDAAN, SURREG, RATIOEST, RTIFREQS, RTILOGIT
Vendor: Research Triangle Institute Batch and interactive IBM S/370 software for sample survey data analysis.

LISREL, PRELIS
Vendor: Scientific Software, Inc. A batch IBM S/370 program that estimates the unknown coefficients of a set of linear structural equations.

MSTAT1
Source: DCRT staff. IBM S/370 batch programs and subroutines for mathematical and statistical analysis.

GLIM (Generalized Linear Interactive Modeling)
Vendor: Numerical Algorithms Group, Inc. An IBM S/370 batch and interactive system for analysis of linear statistical models.

data management system was extensively used at NIH, with an average of 77,600 accesses per month via the IBM® System 370(S/370). The BMDP® package was accessed an average of over 400 times per month.

SSS mainframe statistical support provided maintenance of the system or package and adequate documentation, including NIH computer system changes, system or package updates, and corrections. It also included rapid response to queries about user access to the most used systems and packages. The SSS staff answered over 3,500 calls for software assistance, handling requests for information on job control language, program parameters, and other operating system procedures, as well as assisting in interpretation of results. SSS continues the support of mainframe statistical systems and documentation in response to NIH computer system changes, product updates, and corrections.

Other mainframe software supported by SSS had more limited use. Support for IMSL has included the Convex as well as IBM® S/370 mainframes. There were relatively few sessions for such specialized programs as GLIM and RPART (see Table 3).

While NIH-wide use of statistical software on PC and Macintosh® microcomputers is more difficult to quantify, SSS has continued to expand its support of software on these increasingly popular platforms. The usage of an SSS-negotiated NIH site license for Base SAS®, SAS/STAT®, SAS/GRAPH®, SAS/IML®, and SAS/FSP® has continued to expand. SSS is converting some of the SAS® site-license MS/DOS® products to Windows™. The conversion should be completed by early FY94. SSS also has begun to support SAS® on the SUN® SPARC® station under UNIX®. Several SUN® SPARC® stations are being purchased so that SSS can expand this effort.

Recognizing the importance of teaching the effective use of systems and packages to biomedical researchers and other NIH users, SSS maintained a substantial program of short courses, prepared documentation and held informational talks. Enrollment continued at a high level in the SAS® courses. SSS taught four SAS® courses a total of 25 times to over 300 students through the DCRT training unit. SSS also contracted to have two statistical courses taught through the NIH training center. These two courses were very popular, with 35 students attending.

Future Plans

SSS's high level of support for IBM® S/370 statistical software systems will continue. More statistical software will be supported on the SUN® SPARC® stations, IBM® PC and compatibles, and Mactintoshes. SSS will continue to support the MS/DOS® SAS® site license on the PC. Plans are to convert some of the SAS® site license copies to Windows™. SSS also plans to acquire additional software products for the Windows™ environment. Current plans are to procure a site license for JMP® on the Macintosh® and offer full support. SSS will support SPSS and BMDP® on the SUN® SPARC® station. SSS will in FY94 begin offering LIMDEP on the IBM® S/370.

Due to the success of the SAS®-based statistical courses offered through the NIH training Center, SSS will increase the number of courses in FY94.

OD

Office of the Director

Office of the Director

David Rodbard, M.D., Director
William Risso, Deputy Director

The Office of the Director (OD) provides overall program and management direction for DCRT. The Director, Deputy Director, Associate Directors, Assistant Directors, and Executive Officer work together as the immediate Office of the Director, whose activities in FY93 encompassed such issues as:

- management of the division, including allocation of budget, personnel and other resources
- liaison with NIH/OD and all ICDs
- program evaluation and peer review
- development of new initiatives
- integration of the activities of the offices, laboratories and branches of the division
- development of necessary data to design and acquire the next generation of technology
- interface with regulatory issues and agencies
- support and guidance for computational molecular biology
- division reorganization
- the High-Performance Computing and Communication initiative
- information resources management and strategic planning
- liaison with other Federal agencies.

OD oversees one scientific group, the *Computational Molecular Biology Section* (CMBS), that is charged with supporting and guiding the NIH intramural research community in this area. Section chief Dr. Peter FitzGerald has developed a number of new courses, training manuals and booklets for the users of the GCG sequence analysis package and several genetic databases. He has interacted with literally dozens of scientists from virtually all of the ICDs, providing them assistance and consultation in their analyses of gene sequences. CMBS's Dr. Robert Pearlstein has developed a number of new courses and training materials to assist scientists throughout NIH with molecular modeling.

Three other offices supplement the work of the DCRT laboratories and branches:

The *Office of Information Resources Management* (OIRM) is responsible for coordinating and preparing the DCRT contribution to the NIH IRM Strategic Plan, Tactical Plan and the Environment and Resources Report. The DCRT OIRM will focus on major DCRT procurements, providing planning, oversight and technical guidance. ADP security will continue to evolve and require additional attention.

The *Equal Employment Opportunity Office* (EEO) manages a full EEO program for the division. The office serves as the focal point and advisory for all activities relating to the equal employment opportunities of DCRT employees and applicants. The EEO Officer maintains a close working relationship with the NIH Division of Equal Opportunity and other components concerned with minority and women's issues.

The *Office of Administrative Management* (OAM) provides administrative and managerial support for the work of DCRT. OAM includes the Administrative, Personnel, Financial Management, Project Control and Information Offices, and the DCRT Library.

Computational Molecular Biology Section

Peter C. FitzGerald, Ph.D., Chief

The Computational Molecular Biology Section (CMBS) is the primary group through which DCRT provides support and guidance to NIH intramural scientists in the area of computational molecular biology. The scientific areas addressed by CMBS include the application of computational tools for the collection, analysis and management of primary DNA and protein sequence data, as well as the use of computational techniques to model the chemical structures of proteins, nucleic acids, and other biomolecules. To service the very diverse NIH community, CMBS provides support on a wide variety of computer platforms ranging from personal computers to mainframes. CMBS works closely with

other DCRT labs and branches to provide and maintain a wide variety of resources. Prominent among these intra-DCRT collaborations are those with the Convex System Staff (CFB), members of the Distributed Systems Branch (DSB), and the DCRT Training Unit (CSB).

CMBS staff bring a background in the biological sciences combined with an extensive variety of computer skills to assist NIH researchers in bridging the gap between the very different worlds of "bench research" and computerized data analysis. As well as providing direct user assistance for supported applications, the CMBS has provided scientific consultations to individual NIH scientists interested in: initiating appropriate analysis of data; interpreting the biological significance of computer-based analyses of data; and designing future biological experiments following computer-based analyses of existing data. In its support role the CMBS interacts with NIH scientists from all ICDs, including individuals from both the main NIH Campus in Bethesda and individuals from NIH satellite facilities.

Computational Molecular Biology

During the past year the section has been actively engaged in expanding the variety and quality of programs and facilities available to scientists involved in the area of computational molecular biology. A UNIX® version of the Genetics Computer Group (GCG) sequence analysis software was installed on the DCRT Convex C3830 in early FY93. This implementation represents a major simplification of the GCG software which was previously available on the Convex only under the VAX/VMS™-emulating shell, COVUE®. This software is actively used by over 350 NIH researchers, and constitutes the primary centrally maintained resource available to NIH scientists active in this field. During this year, Dr. FitzGerald continued his efforts to provide training, through several organized courses and invited seminars, and consultation to scientists in the application and use of this software. In addition to support of the GCG software, Dr. FitzGerald continued to maintain current copies of the major DNA and protein sequence databases on the NIH Convex System.

In an effort to provide facilities for DNA and protein sequence analysis on UNIX®-based workstations, Dr. FitzGerald collaborated with members of the Advanced Laboratory Workstation (ALW) staff to put in place a workstation-based version of the GCG sequence analysis software. This software was made available on SUN® workstations as part of the ALW project. CMBS has plans to expand this facility in the coming year to include ALW Silicon Graphic workstations. This latter project will offer scientists (especially those who were part of the original joint purchase of Silicon Graphics® workstations, and who became part of the ALW project) a concrete option for carrying out DNA and protein sequence analysis on these platforms.

With the hope of increasing the ease and speed of access to the major DNA and protein sequence databases, Dr. FitzGerald played a key role in enhancing the NIH *Gopher™* server and UNIX-based client software to provide better integration between molecular biology database searches through *Gopher™* and the sequence analysis tools of the GCG software. Plans are in place to present a wide variety of molecular-biology-related data and user orientation software through the *Gopher™* server in the coming year.

Molecular Modeling and Computational Chemistry

During the past year, CMBS's Dr. Robert Pearlstein has been providing expert support to the NIH intramural community in the area of computational chemistry and molecular modeling. This has included evaluating and procuring software from commercial and public domain sources, implementing and maintaining such software on NIH computers, and supporting these scientific applications via consultation, collaboration, and training. Information about the capabilities, strengths and weaknesses of the various software products and

their requisite hardware platforms was disseminated to NIH scientists interested in purchasing laboratory-based molecular modeling systems tailored to particular research needs.

To provide a general utility for viewing and manipulating three-dimensional protein structures on personal computers, CMBS procured a site license for MacImdad™, a Macintosh®-based product developed at Stanford University. MacImdad™ is a self-contained program for the Macintosh® II that provides manipulable color representations of proteins and other molecules, as well as a compressed version of the entire Brookhaven Protein Databank (PDB). The introduction of this software to the NIH campus, in early FY93, was marked by a seminar and training session presented by Dr. Michael Levitt, the developer of the program. This project has enjoyed considerable success with more than 175 copies of the software being distributed to interested NIH scientists.

As part of CMBS's commitment to facilitating the use of computational chemistry and molecular modeling software, Dr. Pearlstein developed and presented a number of training courses for NIH scientists. These include "Introduction to Molecular Modeling" and "Introduction to Sybyl." These classes have been attended by approximately 200 NIH scientists. In addition, CMBS arranged for a week-long intensive training course on the Quanta®/CHARMm® software package, presented by Dr. Don Kyle of Scios Nova, Inc. The course was given twice: in March and September 1993. Approximately 50 NIH scientists have attended the two sessions. CMBS expects to continue to devote considerable time and effort to user education in coming year.

Dr. Pearlstein has collaborated on a number of scientific research projects throughout this past year. Examples of such projects include:

- structure-activity studies of novel antitumor compounds
- studying the three-dimensional structure-activity of anticataract drugs
- studying the conformational properties of some cyclic dipeptides.

Dr. Pearlstein has been involved in writing an extensive guide to computational chemistry software for NIH scientists, entitled "The SCRC Handbook of Molecular Modeling." This work, as an ongoing project with in-progress editions, is being made available through the SCRC.

CMBS has played a pivotal role in providing expert and technical assistance to the SCRC staff in maintaining their computational chemistry/molecular modeling resources. The SCRC provides scientists with access to a wide variety of computer hardware, software and peripherals, as well as providing an identifiable portal for obtaining access to the extensive resources available within DCRT. A broad range of computational molecular biology and computational chemistry software is accessible through the SCRC, and CMBS expects to continue to play an active role in providing scientific support for these applications.

NIH Gopher™ Server

Having played a major role in the original implementation of the NIH *Gopher™* Server, Dr. FitzGerald became the project leader in FY93 for the continued development, enhancement and maintenance of this very versatile information delivery system. A joint project between CMBS and CFB the *Gopher™* client/server-based information search and retrieval system (developed at the University of Minnesota) has proven to be a major success for the simple and reliable distribution of a wide variety of information.

Following a major reorganization of its menu-based data structures and an enhancement of its search capabilities, the NIH *Gopher™* server now provides access to information on such topics as:

- health-related information and clinical protocols
- NIH grant and contract notices
- molecular biology databases
- images of PDB protein structures
- bibliographic reference data (Current Contents®, REFERENCE UPDATE®)
- weather information

- searchable NIH e-mail directory
- access to more than 1,500 *Gopher™* servers worldwide.

Recent major enhancements to the NIH *Gopher™* server have included:

- the addition of full boolean operators to modify all searches
- an expandable hit list
- the ability to save the contents of a hit list
- full text searches of the local menu structure.

With more than 1,000 accesses a day from a total of more than 5,000 different client machines (both from NIH and around the world), *Gopher™* continues to grow at a rate of approximately 70% per year. In FY93 approximately 500 NIH personnel attended seminars and training classes describing the NIH *Gopher™* Server. The expansion and enhancement of this facility is expected to continue into the coming year, with some major enhancements already in development.

Publications

FitzGerald P. Introduction to GCG – sequence analysis on the NIH Convex system, Parts I and II, DCRT November 1992.

FitzGerald P. User's Guide to Gopher™ at NIH, DCRT October 1992.

FitzGerald P. C., Hartley R.W. Polyethenoadenosine phosphate as a fluorogenic substrate for barnase, *Anal Biochem* 1993; 214(2):544-47.

Bivin D., Kubota S., Pearlstein R., Morales, M. Conformational studies of cyclic dipeptides pertaining to fluorescence measurements, *Proc Nat Acad Sci* 1993; 90:6791.

Pearlstein R. The SCRC handbook of molecular modeling, DCRT October 1993.

Raghavan N., Maina C. V., FitzGerald P. C., Tuan R. S., Slatko B., Ottesen E. A., Nutman T. B. Characterization of a muscle-associated antigen from *Wuchereria bancrofti*, *Exp Parasit*, 1992; 73:379-89.

Yong L., Pearlstein R., Kador P. Studies of aldol reductase molecular modeling inhibitors, *J Med Chem* (in press).

Office of Information Resources Management

Arthur Schultz, Chief

DCRT programs constitute a majority of the IRM activities at NIH. Accordingly, in recognition of the critical importance of IRM to the DCRT program, DCRT established an Information Resources Management (IRM) Office within the Office of the Director, DCRT. In March 1993, Mr. Arthur Schultz was selected as Information Resources Management Officer.

Since its establishment, the DCRT IRM Office has concentrated on the development of DCRT contributions to the overall NIH Strategic Plan. DCRT is the principal support for the information technology needs of the NIH intramural, extramural, and administrative communities. The DCRT planning process has focused on coordinating cross-cutting division initiatives. In response to the new DHHS reporting process, DCRT prepared and submitted three separate IRM planning documents:

- the IRM Strategic Plan: addresses major program goals and information needs and IRM goals and strategies, and follows from a broad DCRT strategic plan
- the IRM Tactical Plan: defines tactical planning assumptions, identifies major IRM initiatives, and provides the status of previous major initiatives
- the IRM Environment and Resources Report: describes current information technology resources, IRM accomplishments, improvements and/or significant changes during the past year.

The preparation of these reports was coordinated by the DCRT IRM Office from material submitted from throughout the division.

Several IRM functions which had been a DCRT responsibility have been transferred to the NIH Office of Information Resources Management (OIRM) which, in 1992, consolidated NIH IRM planning, policy, capacity management, and oversight into a central IRM component:

- the ADP clearance function
- the regulatory interpretation of the Federal Information Resources Management Regulations (FIRMR) and Departmental Guidelines
- the Policy Coordination function responsible for Information Technology Systems budgeting and tactical planning
- Federal Information Processing Standards
- automated inventories
- automated information systems security
- systems reviews
- policy coordination.

The DCRT staff responsible for these functions also were transferred to OIRM.

DCRT has participated actively in the Strategic Planning Advisory Group under the direction of the NIH OIRM. This group is charged with revising the NIH IRM strategic planning process.

In July, DCRT and NIH OIRM jointly recruited an acquisition specialist to serve as a project leader or "Trail Boss" for all phases of the next major systems procurement for DCRT. A Concept of Operations was completed by DCRT in April 1992 which identified DCRT user requirements and categorized them into six basic categories:

- science specific support
- mainframe processing services
- software support/development
- microcomputer and LAN support
- customer service
- networking support.

The procurement will be implemented under a Trail Boss Charter from the General Services Administration (GSA). This partnership with GSA will allow a streamlined acquisition for the next generation of information systems that will serve the scientific computing needs of NIH. An Advisory Council was established in September and the Trail Boss will operate under a draft charter until it is formally approved by GSA. After that procurement has been completed, the Trail Boss will be assigned by the OIRM to other major NIH procurements.

During the coming year an IRM staff will be assembled, probably by reassigning staff from other components of the division. The initial focus will be on expediting ADP procurement and software licenses, and on continuing the planning process begun this year. ADP procurement will remain difficult and will require advanced planning and continual monitoring to ensure that procurements are awarded in time to meet DCRT requirements. DCRT will attempt to alleviate some procurement problems by using Interagency Agreements and by purchasing goods and services against existing government contracts at other agencies. This initiative has resulted in using the NASA Scientific and Engineering Workstation Procurement (SEWP). DCRT participated in the evaluation of SEWP bids and took the lead in organizing the process which resulted in making SEWP available to NIH. DCRT also participates in a Department of Justice ADP Related Services Contract to meet DCRT requirements.

NIH-wide policy and procedure for approving software license agreements is being established. A recent review of several license agreements within DCRT resulted in a recommendation by the Office of General Council to place the responsibility for review and signature with the NIH contracting official responsible for the software acquisition. DCRT is working with the Division of Procurement to expedite the completion of pending software agreements. In the near future DCRT will examine the possibility of the division sponsoring the purchase of site licenses of widely used software.

In the area of planning and management, the division will accept a central role in the IRM process at NIH, especially in that of technical advisor. DCRT will advocate standardization across the NIH for increased interoperability, and will encourage purchases that are designed for operation in a heterogenous environment. The division will also work toward helping NIH achieve a corporate view of

computing in areas of cost recovery, corporate computing architectures and standards, and will continue to advocate and negotiate for adequate resources to build and maintain the strong computational presence required for modern biomedical research.

For many years, DCRT has provided NIH with guidance and support in computer-based assistive technology primarily in providing voice response input devices for individuals unable to communicate with a computer using a keyboard or trackball. DCRT plans to become a more vigorous leader in providing computer-based assistive technology, possibly by assembling a demonstrable personal-computer-based voice response system as a part of the DCRT Scientific Computer Resource Center.

Providing assistive technology to employees with disabilities has received increased emphasis in the private sector because of the Americans with Disabilities Act signed July 26, 1990, and effective at various times ranging from 30 days to 30 years. In general, Public Accommodations should have been in compliance by January 26, 1992.

Many agencies have approached the support of assistive technology at the agency level in order to have a reasonably sized constituency. Another approach adopted by several agencies is to provide this service using contracts with commercial firms offering technologies for facilitating disabled access. DCRT initially will limit direct support to voice response and provide additional support using a contractor.

Equal Employment Opportunity Office

Gloria I. Richardson, EEO Officer

The Equal Employment Opportunity (EEO) Office manages a full EEO program for the division. The office serves as the focal point and advisor for all activities relating to the equal employment opportunities of DCRT employees and applicants. The EEO Officer maintains a close working relationship with the NIH Office of Equal Opportunity and other components concerned with affirmative employment issues.

Upon the annual evaluation of the DCRT Affirmative Employment Program Plan for Minorities and Women by the NIH Office of Equal Employment Opportunity, the plan was cited as reflecting a "positive commitment" to improve the EEO representation and advancement of minorities and women. Of particular note was the "Open Communication" channel: In October, the Division's Director, EEO/Employee Advisory Committee, and Information Office, sponsored its first Town Meeting. The meeting was designed to enhance communication and job performance of the DCRT staff, thus minimizing complaints of discrimination. More than 100 DCRT staff members attended. Over 100 scientific and administrative questions were submitted to employee committee members or placed in the employee suggestion box. Responses have been provided throughout the year in the division's employee newsletter, "Input/Output." A second annual meeting is contemplated.

Additionally, a Career Enhancement Program (CEP) for in-house employee development was implemented and the Co-op program has been highly supported by management personnel.

Nearly 1,500 students have allowed the DCRT to become their "parents"! Through the D.C. Partners In Education Program, the Woodrow Wilson Senior High School student body was officially "adopted" by the division. The adoption ceremony was joyfully celebrated at Wilson in November. This year, we have presented the following student seminars: Computer Career Opportunities; Overview of DCRT and Workplace Tour; and What's Hot in Computing? The implementation of access to NIHEDNET – an educational network and bulletin board system – and the mentoring of the students are next year's goals.

DCRT's EEO Officer and the NIH Federal Women's Program Manager (Office of Equal Opportunity) Ms. Lucretia B. Coffer were nominated and presented with a Special Act Group Award based on their significant contribution to the prevention of sexual harassment at NIH. In January 1993, Ms. Richardson and Ms. Coffer began developing a Train-

the-Trainer Project whereby NIH employees could be given instruction in training techniques and sufficient information on the prevention of sexual harassment subject matter to conduct training at NIH. This initiative, which included the entire planning process, preparation of course materials, and the execution, was accomplished in addition to their regular work assignments.

The DCRT EEO Office presented its first EEO Orientation Session to the division's Stay-In-School and Co-op students in conjunction with the DCRT Personnel Office.

Office of Administrative Management

Marian Dawson, Chief

The Office of Administrative Management (OAM) provides guidance and support on all administrative and business management aspects of the division's programs, advising on the management of resources, the provision of administrative services, program planning and evaluation, and policy and legislative analysis. During FY93, the OAM organization expanded to include two new components from the former Office of Scientific and Technical Communication: the *DCRT Library* the *DCRT Information Office*. This was carried out as part of the major reorganization of DCRT.

During this past year, the Chief of OAM has been actively involved in the implementation of the DCRT reorganization, space planning and relocation resulting from the DCRT reorganization; planning for the relocation of DCRT as part of the NIH Master Plan study and Program Justification Document development; and management of staff during a time of diminishing resources.

The *Administrative Management Section* (AMS), with a staff of eight, headed by Administrative Officer Marlyn Harrison, continues to place all property inventory online in conjunction with an NIH-wide search and reconciliation of property inventories. In addition, the section is working with the Information Systems Branch to bring DCRT branches online with a new NIH Administrative Data Base feature. This feature, Request for Purchase Action (RPA), provides lab/branch staff with the ability to create a preliminary request for goods and services which can be tracked at any time for status.

In reference to the Director's urging that DCRT have a directory and a means of communicating with other DCRT staff with e-mail on Wylbur, Convex, 3COM, and ALW, staff of the AMS worked with staff of the Personal Computing Branch to incorporate e-mail addresses into the telephone listing for all registered e-mail users. This e-mail user i.d. is now a part of DCRT's monthly telephone listing.

To provide an introduction to new staff and a written source of reference for DCRT in general, AMS staff developed the Administrative Handbook. The handbook provides standardized administrative procedures to be followed, is in loose-leaf form (to accommodate revisions), is divided into six indexed sections for quick access, and contains completed examples of all necessary forms. Production of the newsletter continues to address current and anticipated changes in the areas of property, procurement, travel, training, the visiting scientists and fellows program and various other administrative issues. The quarterly newsletter provides procedural assistance and informs DCRT staff of impending changes both at DCRT and NIH levels.

For easier and quicker access to product specifications on GSA Schedule, the AMS installed SOURCE ONE GSA Service. SOURCE ONE is CD-ROM technology available to all DCRT staff, and provides facsimile images of schedules and catalog pages intended to assist the purchaser's search for procurement information with ease, speed, and efficiency.

The *Budget Office*, headed by Mr. Michael Reed, continues to carry out its financial functions for the division, including budget formulation for its two funding mechanisms, and preparation for budget review by the NIH Central Services Budget Review Committee. The office also administers the allocation of available funds within the division, tracks expenditures for each organizational area, and

provides various financial reports, reviews and analyses for DCRT management. The office coordinates financial aspects of proposed cost recovery plans for the ALW project and networking under the NIH Service and Supply Fund. It also responds to requirements for internal control reviews and supplies materials to NIH/DFM for the annual CFO audit.

The *Project Control Office* (PCO), under Ms. JoAnne Higgins, serves as the financial focal point for all DCRT services, performing account and user registration, advising customers on accounting procedures and billing matters, and maintaining DCRT's Project Accounting System, which provides billing information to users and accounting information to the NIH Central Accounting System.

Due to the reorganization of ADAMHA and the establishment of three new Institutes, PCO served as liaison between DCRT, ADAMHA and the new institutes, providing guidance and information relating to usage of, accounting for, and billing of the data processing services at DCRT. PCO initiated and coordinated accounting and systems changes within DCRT to enable a smooth transition of the data processing accounts of former ADAMHA organizations.

PCO also participated in task groups reviewing cost recovery mechanisms and plans for the Advanced Laboratory Workstation project and networking, with the Project Account System being the vehicle used for billing users of the new services. With cost recovery for ALW beginning in October 1993, all ALW users were asked to re-register with the PCO in order to implement cost recovery for this service. PCO collaborated with the Convex systems team to develop a means of registering these users.

PCO, in collaboration with the Program Support Section, CCB, participated with the systems security staff in the Risk Analysis of NIH Computer Application Systems-the Project Account System, providing information concerning PAS assets.

The PCO participated with the CCB systems team in developing and implementing a system that gives the account sponsors and alternate sponsors the ability to register users and request changes to the IBM® mainframe accounts electronically, directly from their PC terminal or workstation, rather than having to complete the forms previously required. The new system, ENTER SPONSOR, also provides verification of the actions through an electronic mailing address back to the sponsor. The PCO contacted all account sponsors and alternates encouraging participation, and continues to register all sponsors and alternates for access to the online registration system.

In FY93, the PCO opened 200 new accounts and registered 2,400 new users to the IBM® mainframe system. In addition, 1,500 new users were registered on the Convex system, 350 new users were registered to the Advanced Laboratory Workstation Project, 122 new users were registered to POP (Post Office Protocol), and 60 new users were registered to the Intel® Supercomputer System. The PCO also completed its annual update of information on over 3,400 IBM® mainframe accounts and 20,000 users.

The DCRT *Personnel Operations Section* advises and assists management in providing and optimally utilizing human resources to accomplish divisional goals, and is responsible for conducting the Personnel Management Program for the division. This includes staffing and recruitment services, compensation and classification, employee benefits programs, retirement, training, performance appraisal, awards and incentive programs, employee relations, conduct and ethics.

This year, the office was influential in several key initiatives such as the reorganization of the entire division. The Personnel Office was an integral player in the intensive review of DCRT programs, wrote position descriptions and numerous internal vacancy announcements for new positions, and determined position management strategies and staffing patterns for the new organizations.

The Personnel Office developed the DCRT Career Enhancement Program for which several employees from the CCB Operations Section were selected for professional positions within the division. The personnel staff wrote the formal training plans for the participants and coordinated with the supervisors to ensure effective implementation of the

program. The staff also assisted in organizing the Third Annual Division Award Ceremony as well as in writing many award nominations.

The reduction in FTEs now in effect at the NIH is a direct result of the federal government's effort to streamline agencies and programs. The DCRT personnel office was able to significantly augment staff by utilizing student employment programs such as the Stay-in-School program, Co-operative Education Program and the Federal Junior Fellowship Program, with special attention to attracting minority employees. DCRT participated in a joint Partners in Education ceremony with the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) at Woodrow Wilson High School. The partnership agreement formalizes DCRT's commitment to attracting and recruiting talented young people with diverse backgrounds.

The personnel staff provided an increasing variety of assistance and information on employee benefits and services including the first open season for federal life insurance in 8 years. The staff also greatly facilitated the expansion of the Alternative Work Schedule/Work Place Program throughout the division. The office created and distributed two valuable booklets entitled "The DCRT Awards Handbook" and "The Personnel Help Book," in addition to writing over five administrative policy and procedure statements for the division.

The personnel staff successfully completed the arduous task of automating over 300 position descriptions and evaluation statements. This information, now readily available on diskette, allows for easier access and greater efficiency with respect to updating and modifying such documents.

The staff of the Personnel Office has been active in tracking the training of DCRT's supervisory personnel. We have met with each supervisor, on an individual basis, to determine their training needs for the future. The staff coordinated a two-day Supervisory Training and EEO Seminar for over 60 supervisors, leads and project officers. The staff also organized ethics training for supervisors in July 1993.

This year, the DCRT Personnel Office was reviewed with regard to internal controls, procedures and practices in the area of personnel administration. During the course of the review, managers consistently expressed a high degree of satisfaction with the level of knowledge and responsiveness of Personnel Office staff members.

Lastly, the personnel staff has attended specialized training courses, seminars and professional conferences to sharpen skills and keep abreast of new developments in Human Resources. Conferences attended include the NIH Human Resources Management Conference in June 1993, the International Personnel Management Association, Montgomery County Chapter's Spring Conference in April 1993, and Eastern Regional Conference in Harrisburg, Pennsylvania in June 1993.

The *DCRT Library*, under Chief Ellen Chu, provides information resources to NIH staff in computer science, mathematics, and statistics, along with computer applications in biomedical sciences, engineering, information science, and management. Library staff includes two librarians and two part-time students. As an active participant in the work of the division, staff pursue collaborative testing of new computer applications for libraries, evaluation and implementation of electronic publications or information systems for NIH-wide access, and use of a full range of telecommunication-based facilities and internetworking to communicate with their clientele and other libraries.

Major innovations this year involved installation of an integrated library application package and introduction of soundproofing and movable shelving. The library initialized the Scientific and Technical Information Library Automation System (STILAS), a UNIX®-based system, in March. In addition to upgrading STILAS to a new release in July, library staff worked with Computer Facilities Branch (CFB) staff to test the new system in an Advanced Laboratory Workstation (ALW) configuration. The implementation schedule has been ambitious, with migration of catalog, circulation, serials, and acquisitions data and transactions from three different application packages into STILAS modules during installation and upgrade phases. The anticipated gain will be

consolidated lookup access to all of these operations by every staff member. Library users began remote access to the online catalog via the NIH *Gopher™* during the summer. In August, the library shut down to install movable, compact shelving. This new shelving increases capacity 30%, providing some relief to the overcrowding. Soundproofing along the wall shared with the conference room reduces noise at the side of the library where the photocopier is located.

Library staff continued collaborations with the Distributed Systems Branch (DSB), providing a range of information services via PUBnet, the public network established for NIH 3Com, Lan Manager, and Appletalk® networks. Several CD ROM products were mounted for user evaluations. Based on responses, two publications from Datapro, *Communications Analyst* and *Computer Systems Analyst* were acquired for NIH-wide use in the latter part of the year. In addition to *Computer Select*, the library added *Physician's Desk Reference/Merck Manual* and *SOS/Applications* to PUBnet offerings. Library staff began evaluating the technical feasibility of providing campus-wide access to a full-text journal with images on CD ROM, the *Journal of Biological Chemistry*. The library renewed its license to the *Current Index to Statistics* database for the NIH *Gopher™* service on the Convex.

As the DCRT representative to the NIH Library Advisory Committee, the library Chief worked with various DCRT components to facilitate collaborations with the NIH Library in providing NIH-wide services. An early effort was coordination and cooperation in support of bibliographic reference database software packages: selection, training, support. NIH scientists use these applications to manage their bibliographic reference files and to transfer their citations for the NIH *Scientific Directory/Annual Bibliography*. A second endeavor – involving DCRT, the NIH Library, and the National Library of Medicine – is support of NIH end-user access to *MEDLINE* via GRATEFUL MED® on the Internet. Finally, periodic meetings involving NIH Library, CFB, and *Gopher™* staff have enhanced communications regarding NIH *Gopher™* and NIH Library campus information system plans. Advance notice of plans coordinates procurement to avoid duplication and also enables NIH Library staff to anticipate user reference queries. Investigations continued in search of a multiplatform CD ROM networking solution.

This year, 163 new users registered to borrow, an increase of 14% over last year, with 70% from other parts of NIH.

Next year, the library looks forward to consolidation after fundamental changes in daily operations with the new STILAS and movable shelving systems. There are plans to install additional STILAS modules which will provide staff and users access to remote and local databases using the same search engine as the online catalog. The library anticipates a resurgence of DCRT staff use of the library, as reorganized units pursue new projects. Purchase of remote journal check-in service will provide staff support in a labor intensive area. Developmental work with electronic publishing, campus-wide information systems, and multiplatform networking of CD ROMs will proceed as resources permit.

The DCRT *Information Office*, under Information Officer Raymond Fleming, is responsible for communicating the advantages of computer research and technology to NIH administrative and scientific communities; electronic and print media, including the trade press; users of the DCRT Central Computer Utility; and the general public. Equally important is the responsibility for helping to build and maintain communication within DCRT. To accomplish these objectives, a staff of three public affairs specialists are involved in the following functions:

- publications and articles
- special events
- support for the Office of the Director and the larger NIH community
- corporate identity and image development
- media liaison
- public inquiries.

Office writers produced the *1992 Director's Report*, a draft of the 1993 report, a much-awaited

update of *Computing Resources* and the *NIH Directory of Image Processing Facilities.*

A major feature story was written for the NIH Record on the DCRT reorganization; this story formed the basis for a poster at the NIH Research Festival. Other articles featured in the Record and the new NIH Catalyst included a DCRT/NIAAA collaboration, cluster computing, and the new Laboratory of Structural Biology; numerous shorter stories and photo/captions related to division activities were also sent.

At the suggestion of Information Office staff, the Record began accepting articles electronically, thus saving considerable time and energy. Several issues of the employee newsletter, Input/Output, were produced; training stories and many other articles were edited for the Record and PCBriefs, and a press release on the BOSS (Best of Open Systems Solutions) award was written and distributed. Staff also contributed updates to the Scientific Directory/Annual Bibliography (SD/AB), international and audiovisual reports, the NIH Almanac, the NIH Calendar of Events and Meetings, and the National Research Council Associate Program Booklet.

Fiscal Year 1993 saw the Information Office become heavily involved in graphic arts and photography. Over 30 photo assignments were carried out, and numerous posters, tent cards, fliers, and other items were developed on the desktop or in conjunction with Medical Arts. Special displays produced by the office included:

- the DCRT reorganization
- the Network Systems Branch
- the Computational Bioscience and Engineering Laboratory
- the Scientific Computing Resource Center
- a new telephone file card with all major DCRT numbers
- the Best of Open Systems Solution (BOSS) Award
- Howard University's Dr. Percy Julian Award to Dr. Bernard Brooks
- the DCRT Awards Ceremony, holiday party, and picnic
- several DCRT scientists and selected research papers, including Drs. V. Adrian Parsegian and Peter Munson, along with several other DCRT scientists contributing to molecular biology.

The office also placed bulletin boards for DCRT notices in stairwells and other locations, developed a display case for the Pratt Conference Room, and had large flannelboards mounted throughout the Bldg. 12 complex to display posters from the NIH Research Festival.

Special events were a major part of the office's FY93 activities. The staff was significantly involved in the planning, execution, and followup of DCRT's first-ever Town Meeting. Staff members organized the highly successful division picnic and the holiday party, assisted with the annual DCRT awards ceremony, and played a significant role in organizing a joint DCRT/NCBI Research Festival "Workshop on Computing in Molecular Biology," which featured eight exhibits and demonstrations and six presentations (DCRT staff also presented over 15 posters over a 2-day session). In addition, planning was begun for the celebration of DCRT's 30th Anniversary in 1994.

The office was also called upon to assist in such events as:

- the High Performance Computing in Chemistry meeting
- the Art and Science of Experimental Design seminar series
- talks for students from Woodrow Wilson High School, DCRT's "Partners in Education" adopted school
- talks for the heads of industrial research from several Fortune 500 companies.
- tours for Japanese scientists, given in Japanese by Mr. Fred Yamada.

One area of significantly increased activity for the office was support for the DCRT Director's initiatives. Information office staff provided background information, slides, publication packets, and other materials for presentations before:

- the NIH Deputy Director for Intramural Research
- the Central Services Budget Review committee
- the Information Resources Management council
- scientists interested in molecular modeling

- the Director of the National Center for Human Genome Research
- incoming postdoctoral fellows
- NIDA, NIA, NIEHS, EPA, and NIST.

In addition, office staff attended a congressional hearing on High Performance Computing and Communication, and tracked several pieces of legislation, speeches, and testimony.

The office turned its attention to several major projects affecting the NIH community at large. Information office staff played leading roles in coordinating and implementing a new electronic submission process for the NIH Scientific Directory/ Annual Bibliography (SD/AB), and wrote an SD/AB software user's guide. These contributions should result in a faster and more efficient publication process for the NIH Office of Communications. The office also set up an e-mail group for NIH Information Officers, and gave a presentation on its use.

Media activity continued to be an increasing office priority. Assistance was provided to Federal Computer Week and Government Computer News, U.S. News & World Report, PC World, MacWeek, Open Systems Today, ComputerWorld, Healthcare Competition Week and Today's Chemist. In addition, DCRT staff experts were featured in a videotaped segment for the program "Mac Today," and in an internationally broadcast business program involving new software.

Two staff members served on the DCRT Employee Advisory Committee, becoming involved in its many initiatives. Staff also coordinated regularly with the Scientific Computing Resource Center. Other areas to which the office contributed included disaster recovery and the "Partners in Education" and Marriott "Bridges" programs.

Finally, the Office continued its tradition of service to its external and internal public through its many daily information activities. Staff members ably handled information, publication, and photo requests ranging from several to a score per week, tracked division-submitted scientific papers, and regularly issued computer clips from major daily newspapers. Eight Freedom of Information requests were processed, and a major effort at reorganizing the office's files paid large benefits. To keep pace with computer-based advances in the public information arena, the staff trained in Advanced Macintosh®, *Gopher™*, and KaleidaGraph™. Other training included telephone techniques, sexual harassment prevention, and the Privacy Act.

ACRONYMS

ABI	Applied Biosystems, Inc.
ACH	Automated Clearing House
ADAMHA	Alcohol, Drug Abuse, and Mental Health Administration
ADB	Administrative Data Base
ADBIS	Administrative Data Base Information System
ADP	Automatic Data Processing
AECG	Ambulatory Electrocardiography
AJCC	American Joint Committee on Cancer
ALW	Advanced Laboratory Workstation
AMS	Administrative Management Section
ANCOVA	Analysis of Covariance
ANNs	Artificial Neural Networks
AR	Autoregressive
ARAP	Appletalk® Remote Access Protocol
ARMA	Autoregressive Moving Average
ASPS	Applied Systems Programming Section
ATM	Asynchronous Transfer Mode
BCS	Biostatistical Consulting Section
BEIP	Biomedical Engineering and Instrumentation Branch
BFSB	Biometry and Field Studies Branch
BIMAS	BioInformatics and Molecular Analysis Section
BOSS	Best of Open Systems Solutions
BPB	Biological Psychiatry Branch
bR	Bacteriorhodopsin
BRB	Bone Research Branch
BRMUG	Biomedical Research Macintosh® Users Group
CAP	Cluster Analysis Program
CART	Cartesian and Regression Tree Classification
CAS	Central Accounting System
CAS	Clinical Applications Section
CASE	Computer-Aided Software Engineering
CB	Cardiology Branch
CBEL	Computational Bioscience and Engineering Laboratory
CC	Clinical Center
CD-ROM	Compact Disk-Read Only Memory
CEP	Career Enhancement Program
CERT	Computer Emergency Response Team
CFB	Computing Facilities Branch
CFO	Chief Financial Officer
CHIPS	Child Health Information System
CIU	Clinical Information Utility
CMBS	Computational Molecular Biology Section

CMI	Computerized Microfilm Index System
CMS	Capacity Management Staff
CPU	Central Processing Unit
CRADA	Cooperative Research and Development Agreement
CRISP	Computer Retrieval of Information on Scientific Projects
CSB	Customer Services Branch
CSTO	Computing Systems Technology Office
CT	Computed Tomography
CURE	Campus User Research Exchange
DARPA	Defense Advanced Research Projects Agency
DBAS	Data Base Applications Section
DBIS	Data Base Information Section
DBSS	Database Systems Section
DBTG	Database Technology Group
DCBDC	Division of Cancer Biology, Diagnosis, and Centers
DCPC	Division of Cancer Prevention and Control
DCRT	Division of Computer Research and Technology
DCT	Division of Cancer Treatment
DFM	Division of Financial Management
DFT	Discrete Fourier Transform
DMS	Data Management System
DNA	Deoxyribonucleic Acid
DNM	Department of Nuclear Medicine
DOS	Disk Operating System
DPM	Department of Personnel Management
DRD	Diagnostic Radiology Department
DSA	Digital Subtraction Angiography
DSB	Distributed Systems Branch
DSS	Distributed Systems Section
ECG	Electrocardiogram
ECL	Electrocardiogram Criteria Language
EEG	Electroencephalogram
EEO	Equal Employment Opportunity Office
EFT	Electronic Funds Transfer
EIB	Experimental Immunology Branch
EM	Expectation Maximization
EMBL	European Molecular Biology Laboratory
EMT	Environmental Management Tool
EPA	Environmental Protection Agency
ERIs	Electron Repulsion Integrals
ESDS	Enterprise Systems Development Section
ESR	Electron Spin Resonance
ETS	Enterprise Technologies Section
FC/ADA	Flow Cytometry/Advanced Data Analysis
FCC	Federal Computer Conference

FCCSET	Federal Coordinating Committee for Science, Engineering and Technology
FDA	Food and Drug Administration
FDDI	Fiber Distributed Data Interface
FDF	Fast Data Finder
FFT	Fast Fourier Transform
FHA	Filamentous Hemagglutinin
FOV	Field of View
FTEs	Full Time Equivalents
FTP	File Transfer Protocol
FTS	Federal Telecommunications Service
GA	Genetic Algorithm
GCG	Genetics Computer Group
GRC	Gerontology Research Center
GSA	General Services Administration
GUI	Graphical User Interface
Hb	Hemoglobin
HFSS	Hierarchical File Storage System
HPCC	High Performance Computing and Communications
HPCS	High Performance Computing Section
HPSCS	High Performance Scientific Computing Section
I/O	Input/Output
ICDs	Institutes, Centers and Divisions
IF	Intermediate Filament
IMPAC	Information for Management, Planning, Analysis, and Coordination
IPRS	Image Processing Research Section
IR	Division of Intramural Research
IRM	Information Resources Management
IRP	Intramural Research Program
ISB	Information Systems Branch
ITC	Image Technology Center
LANs	Local Area Networks
LBM	Laboratory of Biochemistry and Metabolism
LCB	Laboratory of Chemical Biology
LCB	Laboratory of Cell Biology
LCE	Laboratory of Comparative Ethology
LDACS	Laboratory Data Acquisition and Control System
LDRR	Laboratory of Diagnostic Radiology Research
LMB	Laboratory of Molecular Biology
lod	log of the odds ratio
LP	Laboratory of Pathology
LPP	Laboratory of Psychology and Psychopathology
LSB	Laboratory of Structural Biology
LTIB	Laboratory of Tumor Immunology and Biology
LTPB	Laboratory of Theoretical and Physical Biology
MbCO	Carboxymyoglobin

MDB	Molecular Diseases Branch
MEM	Maximum Entropy Method
MFLOPS	Million Floating Point Operations per Second
MGS	Molecular Graphics and Simulation Section
MHC	Major Histocompatibility Complex
MIMD	Multiple Instruction Stream, Multiple Data Stream
MIPS	Million Instructions per Second
MIS	Medical Information System
ML	Maximum Likelihood
MRI	Magnetic Resonance Imaging
MRIPS	Multimodality Research Image Processing System
MS	Microsoft®
MSS	Management System Storage
MVS	Multiple Virtual Storage
NASA	National Aeronautics and Space Administration
NCBI	National Center for Biotechnology Information
NCHGR	National Center for Human Genome Research
NCI	National Cancer Institute
NCRR	National Center for Research Resources
NEI	National Eye Institute
NHLBI	National Heart, Lung and Blood Institute
NIA	National Institute on Aging
NIAAA	National Institute on Alcohol Abuse and Alcoholism
NIAMS	National Institute of Arthritis and Musculoskeletal and Skin Diseases
NICHD	National Institute of Child Health and Human Development
NIDA	National Institute on Drug Abuse
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases
NIEHS	National Institute of Environmental Health Sciences
NIH/OD	National Institutes of Health/Office of the Director
NINDS	National Institute of Neurological Disorders and Stroke
NINR	National Institute for Nursing Research
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
NQS	Network Queuing System
NSB	Network Systems Branch
NTAS	New Technology Analysis Section
NTSC	National Television System Committee
OAD	Office of the Associate Director
OAM	Office of Administrative Management
OCB	Office of Computational Biosciences
OCRS	Office of Computing Resources and Services
OGCS	Ophthalmic Genetics and Clinical Services Branch
OIRM	Office of Information Resources Management
OSTP	Office of Science and Technology Policy
PBQS	Parallel Batch Queuing System

PC	Personal Computer
PCA	Principal Component Analysis
PCO	Project Control Office
PCR	Polymerase Chain Reaction
PDB	Brookhaven Protein Databank
pdf	Probability Density Function
PEGs	Polyethylene Glycols
PET	Positron Emission Tomography
PLS	Partial Least Squares
PMT	Photomultiplier Tube
POP	Post Office Protocol
PPP	Point-to-Point Protocol
PSL	Physical Sciences Laboratory
PUBnet	NIH Public Network
QM	Quantum Mechanical
QM-MM	Quantum Mechanics-Molecular Mechanics
QMF	Query Management Facility
RAM	Random Access Memory
rCBF	Regional Cerebral Blood Flow
RCOMM	Remote File Access and Communication System
ROB	Radiation Oncology Branch
ROCs	Receiver-Operating Curves
RPA	Request for Purchase Action
RPC	Remote Procedure Call
RRV	RR variability
RSB	Radiation Safety Branch
RT	Reverse Transcriptase
SB	Stroke Branch
SCF	Self-Consistent Field
SCRC	Scientific Computing Resource Center
SCSI	Small Computer System Interface
SD/AB	Scientific Directory/Annual Bibliography
SEWP	Scientific and Engineering Workstation Procurement
SGHM	Slow Growth Homology Modeling
SIU	System International Units
SMTP	Simple Mail Transport Protocol
SOMS	Systems Operations Management Section
SPECT	Single Photon Emission Computer Tomography
SPM	Statistical Parametric Mapping
SRM	System Resource Manager
SSFAS	Service and Supply Fund Activity System
SSS	Statistical Support Staff
STILAS	Scientific and Technical Information Library Automation System
SVD	Singular Value Decomposition
TCP/IP	Transmission Control Protocol/Internet Protocol

TIM	Triose Phosphate Isomerase
TIO	Technical Information Office
TLCs	Technical Local Area Network Coordinators
UPSs	Uninterruptable Power Supplies
URC	User Resource Center
USP	United States Pharmacopoeial
WAN	Wide Area Network

INDEX

Disc risions of
app ongress since
19 shall, on the
gr in, religion,
ha from participation
in subjected to
d or activity.
T ter Research and
Technology compliance with
these laws.

DCRT

Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland, 20892